

Signals and Systems

Dr. Vikram Gadre

January 25, 2016

Contents

I. Natural Domain	1
1. Introduction to Abstractions	3
1.1. Module 1: Lectures 1 - 3	3
1.1.1. Introduction	3
1.1.2. Abstractions	3
1.1.3. Definitions	4
1.1.4. Properties of Systems	6
1.2. Module 1: Lecture 4	
System Properties: An Illustration	7
1.2.1. Introduction	7
1.2.2. RC Circuit	8
1.2.3. Introduction	9
1.2.4. RC Circuit	9
1.3. Module 1: Lecture 5	
Challenging Problems	10
1.3.1. Introduction	10
1.3.2. Two systems	10
1.4. Module 1: Lecture 6	
Impulse Response: A Building Block for Continuous Functions	11
1.4.1. Introduction	11
1.4.2. Narrow pulses	11
1.4.3. Sifting property	12
1.4.4. Stitching together a function with narrow pulses	12
1.5. Lecture 7: Impulse Response and Properties of Systems	13
1.5.1. The Continuous Variable Delta Function	13
1.5.2. Unit Impulse Response	14
1.6. Lecture 8: Relating Unit Step and Unit Impulse	16
1.6.1. Impulse Response of for an RC circuit	16
1.7. Lecture 9: Unit Impulse Response of the RC circuit	19
1.7.1. Taking the derivative of the input signal for an LSI system	19
1.8. Lecture 10: Introduction to Discrete Systems	21
1.8.1. Introduction	21
1.8.2. Definition	21
1.8.3. Representation	21
1.8.4. Examples of a Discrete Time System	22
1.9. Lecture 11: Properties of Discrete Systems.	22
1.9.1. Additivity	22
1.9.2. Introduction	22
1.9.3. Definition	23
1.9.4. Proving Additivity	23
1.9.5. The Tax Example	23

Contents

1.9.6.	Homogeneity and Shift Invariance.	24
1.0.7.	Introduction	24
1.0.8.	Definition	24
1.0.9.	A Homogeneous System	24
1.0.10.	A Shift Invariant System	24
1.0.11.	Causality and Stability.	26
1.0.12.	Introduction	26
1.0.13.	Causality	26
1.0.14.	Stability	26
1.1.	Lecture 12: Impulse Response and Convolution	27
1.1.1.	Introduction	27
1.1.2.	A Linear Shift Invariant System	27
1.1.3.	Theorem	29
2. The Impulse Response for Discrete LSI Systems		31
2.0.1.	Introduction	31
2.0.2.	Proof of the Theorem	31
2.0.3.	Introduction to Convolution	32
3. Discrete Convolution		33
3.0.1.	Introduction	33
3.0.2.	Definition	33
3.0.3.	Procedure for Convolution	33
3.0.4.	Applications	34
3.0.5.	Examples on Convolution	34
3.0.6.	Properties of Convolution	35
3.1.	Lecture 13: Causality of LSI systems	35
3.1.1.	Associativity and Commutativity of LSI Systems	35
3.1.2.	Causality of a Linear Shift Invariant System from its Impulse Response	37
3.2.	Lecture 14: Conditions for Stability	38
3.2.1.	Stability of LSI system from it's impulse response: Discrete variable.	38
3.2.2.	Stability of LSI system from it's impulse response: Continuous variable.	39
3.2.3.	Necessity of the condition for stability	40
3.3.	Lecture 15: Introduction to Convolution	41
3.3.1.	Input Output Relationship in LSI Systems: the Impulse Response	41
3.3.2.	Getting the output for a discrete system	42
3.3.3.	Getting the Output for the General Case: the Train Platform Analogy	44
3.4.	Lecture 16: The Convolution Operator	45
3.4.1.	Introduction	45
3.4.2.	Convolution of Continuous Signals	45
3.4.3.	Convolution of continuous signals – Example	46
3.5.	Lecture 17: Properties of the Convolution Operator	46
3.5.1.	Introduction	46
3.5.2.	Commutativity	47
3.5.3.	Associativity	48
3.5.4.	Consequences of Commutativity and Associativity of Convolution	49

3.6.	Lecture 18: Conclusion of Module 1 & Summary	50
3.6.1.	Introduction	50
3.6.2.	Typical System Description of LSI System	50
3.6.3.	Summary of Module One	51
II. Frequency Domain		53
4. Sinusoids and LSI Systems		55
4.1.	Module 2: Lecture 1	
Sinusoidal Signals and their Properties	55	
4.1.1.	The Ideal Voltage generator	55
4.1.2.	Properties of sinusoids	56
4.1.3.	Sinusoids as Rotating Complex Numbers	57
4.2.	Module 2: Lecture 2	
Sinusoidal Input to LSI Systems	59	
4.2.1.	Understanding Phase Change In Sinusoids and Phasors	59
4.2.2.	Phasor Input to LSI System	60
4.2.3.	Sinusoid Input to LSI System	60
4.2.4.	Periodic Input to Shift Invariant Systems	61
4.3.	Module 2: Lecture 3	
Signals and Vectors	62	
4.3.1.	Relations between Signals and Vectors	62
4.3.2.	Sinusoids as basis vectors	64
4.4.	Module 2: Lecture 4	
Decomposition of Signals	65	
4.4.1.	Introduction	65
4.4.2.	Orthogonal vectors at a particular frequency	65
4.4.3.	Projection of Vectors	67
4.4.4.	Decomposition of square wave	68
4.4.5.	Examples of Decomposition	69
4.5.	Module 2: Lecture 5	
Fourier Series Decomposition and its Applications	71	
4.5.1.	Introduction	71
4.5.2.	Complex Exponential Fourier Series Decomposition	71
4.5.3.	Periodic Input to a Simple RC Circuit	73
4.5.4.	Periodic Inputs to a General Linear Shift Invariant System	74
4.6.	Conclusion	74
5. The Fourier Transform		75
5.1.	Module 2: Lecture 6	
The Fourier Transform	75	
5.1.1.	Introduction	75
5.1.2.	Recapitulation	75
5.1.3.	Interpretation of $H(\Omega)$ as a Dot Product	75
5.1.4.	Inverse for the Fourier Transform	76
5.1.5.	Orthogonality of the Complex Exponentials	78
5.1.6.	Implications	79

Contents

5.2. Module 2: Lecture 7	
Fourier Transforms properties and Fourier series example	79
5.2.1. Introduction	79
5.2.2. Fourier transform of a Rectangular Pulse	79
5.2.3. Properties of Fourier Transform	81
5.2.4. Time Shift	82
5.2.5. Conclusion	84
5.3. Module 2: Lecture 8	
Convolution Property of the Fourier Transform	84
5.3.1. Introduction	84
5.3.2. Convolution Theorem	84
5.3.3. What is a transform?	86
5.3.4. Convolution or Multiplication	86
5.3.5. The Principle of Duality	88
5.4. Module 2: Lecture 9	
Multiplication Theorem and Parseval's Theorem	89
5.4.1. Introduction	89
5.4.2. Multiplication	89
5.5. Module 2: Lecture 10	
Spectral Density	91
5.5.1. Introduction	91
5.5.2. Invariance of Inner Product on Change of Basis	92
5.5.3. Energy of a Signal	92
5.5.4. The Differentiation Property of the Fourier Transform	94
III. Sampling	97
6. chapter	99
6.1. Lecture 1: Introduction: The Basic Theme of Module Three	99
6.1.1. Introduction	99
6.1.2. Recap of EE210.1x	99
6.1.3. Sampling	99
6.1.4. A Priori Information	101
6.2. Lecture 2: Non-uniqueness due to sampling	102
6.2.1. Introduction	102
6.2.2. Non-uniqueness due to lack of a-priori Information	103
6.2.3. Sampling Sinusoid	104
6.3. Lecture 3: Concept and expression of Impostor Terms	106
6.3.1. Introduction	106
6.3.2. Non-Uniqueness of Samples	106
6.4. Lecture 4: Understanding the Impostor Terms	107
6.4.1. Introduction	107
6.4.2. Adding the sinusoid terms	108
6.4.3. An Alternate Path - By Sampling	110
6.4.4. Another Way of Sampling	112
6.4.5. Conclusions	113
6.5. Lecture 5: Sampling of signals having Fourier Transform	114
6.5.1. Sampling	114

6.5.2. Signals which have and do not have a Fourier transform	116
6.5.3. Linearity of sampling	118
6.6. Lecture 6: Sampling Theorem	118
6.6.1. Introduction	118
6.6.2. Sampling Theorem	119
6.6.3. Proof	119
6.7. Lecture 7: Reconstructor: Ideal and practical	120
6.7.1. Introduction	120
6.7.2. Frequency response for Ideal filter for reconstruction	120
6.7.3. Impulse response of ideal filter for reconstruction	122
6.7.4. Problems with the ideal reconstructor	122
6.8. Lecture 8: An example of reconstructor	125
6.8.1. Introduction	125
6.8.2. An Example: RC circuit	125
6.8.3. Properties of a realizable reconstructor	127
6.9. Lecture 9: Nonideal Sampling with Pulses	127
6.9.1. Introduction	127
6.9.2. The Pulse Train	127
6.9.3. Sampling with the Pulse Train	129
6.9.4. Graphical Representation of the Fourier Transform of the Sampled Signal	130
6.10. Lecture 10: Nyquist Sampling Theorem: Proof and Application	132
6.10.1. Introduction	132
6.10.2. General form of the Nyquist-Shannon Sampling Theorem	132
6.11. Lecture 11: Sample and Hold	134
6.11.1. Introduction	134
6.11.2. Interpolation	134
6.11.3. Frequency domain analysis	136
6.12. Lecture 12: Sampling followed by Linear Interpolation	139
6.12.1. Introduction	139
6.12.2. Linear Interpolation	139
6.12.3. Signal Representation of Linear Interpolation	140
6.12.4. Significance in frequency domain	141
6.13. Lecture 13: Applications of Discrete Time Processing	142
6.13.1. Introduction	142
6.13.2. Why Digital Systems over Continuous Systems?	142
6.13.3. A Digital System	143
6.13.4. A Speech Processing System	143
6.13.5. An Overview of a Successive Approximation ADC	145
6.13.6. An Overview of a DAC	146
6.13.7. Frequency Normalisation	147
6.14. Lecture 14: Discrete Time Fourier Transform	149
6.14.1. Introduction	149
6.14.2. Discrete Time Fourier Transform	150
6.14.3. Processing the Discretized Signal	151
6.14.4. Pondering About DTFT	153
6.15. Lecture 15: Discrete time signal processing system	153
6.15.1. Introduction	153
6.15.2. General Discrete-time processing	153

Contents

6.15.3. Going from continuous-time to discrete-time signal processing	154
6.15.4. Graphical view of restriction and normalisation of $H(\Omega)$	155
6.15.5. Example of Constructing a Discrete-Time System Equivalent to an RC Filter	156
6.15.6. Significance of certain frequencies	157
6.15.7. Description of the equivalent discrete-time RC filtering process . .	158
6.16. Lecture 16: Conclusion of Module 3	159
6.16.1. Introduction	159
6.16.2. Recap	159
6.16.3. Construction of equivalent discrete system	159
6.16.4. Inverse Discrete Time Fourier Transform	161
6.16.5. Exact Impulse Response	161
6.16.6. Conclusion of this module	162
IV. Laplace and Z-Transforms	163
7. chapter	165
7.1. Module 4: Lecture 1	
Introduction to Z Transform	165
7.1.1. Introduction	165
7.1.2. Introduction of the Z Transform	166
7.2. Module 4: Lecture 2	
Defining the Laplace Transform and the Z Transform	167
7.2.1. Introduction	167
7.2.2. Laplace Transform	167
7.2.3. Significance of the Region of Convergence	169
7.2.4. Z Transform Equivalent Summary	170
7.2.5. Conclusion	170
7.3. Module 4: Lecture 3	
Examples of Laplace and Z Transforms	171
7.3.1. Introduction	171
7.3.2. Laplace Transform Example	171
7.3.3. Z Transform Example	172
7.3.4. Conclusion	172
7.4. Module 4: Lecture 4	
Properties of Laplace and Z Transforms	173
7.4.1. Introduction	173
7.4.2. Linearity	173
7.4.3. Shifting of Independent Variable	175
7.4.4. Modulation Property	176
7.4.5. Introduction To Convolution Property	177
7.4.6. Conclusion	178
7.5. Module 4: Lecture 5	
Region of Convergence(ROC), Convolution Property and System Function	178
7.5.1. Introduction	178
7.5.2. ROC	179
7.5.3. New way to characterize the LSI systems	179
7.5.4. Convolution property for the z-transform	179

7.5.5. Examples	181
7.5.6. System Function	183
7.5.7. Importance of Region of Convergence	183
7.5.8. System Function as Ratio of transforms	183
7.5.9. A Neat Way to Convolve	184
7.5.10. Conclusion	184
7.6. Module 4: Lecture 6	
Dealing with Poly Ex Terms and Partial Fraction Expansion	184
7.6.1. Introduction	184
7.6.2. Polynomial exponential function	185
7.6.3. Differentiation property for Laplace transform	185
7.6.4. Region of Convergence	186
7.6.5. Partial Fraction Expansion	187
7.6.6. Conclusion	188
7.7. Module 4: Lecture 7	
Dual of the Differentiation Property and Rational Laplace Transforms . .	188
7.7.1. Introduction	188
7.7.2. Rational Laplace Transforms	188
7.7.3. Inverse of the Laplace Transform	189
7.7.4. Irrational Systems	193
7.7.5. Conclusion	194
7.8. Module 4: Lecture 8	
System Analysis Using Laplace Transform	194
7.8.1. Introduction	194
7.8.2. Poles and Zeroes	194
7.8.3. Singularity Functions	195
7.8.4. RC Circuit	196
7.9. Module 4: Lecture 9	
Region of Convergence(ROC), Convolution Property and System Function	199
7.9.1. Differentiation property of <i>z-transform</i>	199
7.9.2. How to handle rational <i>z-transforms</i> ?	200
7.9.3. <i>z-transforms</i> with repeated factors	202
7.9.4. Differentiation w.r.t. 'z' : Time Shifting in the natural domain .	203
7.9.5. System Function as a ratio	204
7.9.6. Zeroes and Poles of the system	204
7.9.7. General process for inversion of rational <i>z-transform</i>	205
7.10. Module 4: Lecture 10	
Looking at rational Laplace transform and z transform together And	
Determining Causality of a LTI system from the system function	206
7.10.1. Introduction	206
7.10.2. System function	206
7.10.3. Looking at rational Laplace transform and z transform together .	207
7.10.4. Illustration	208
7.10.5. Determining the causality of a LTI system from its system function	210
7.11. Module 4: Lecture 11	
Determining system stability using Laplace transforms	212
7.11.1. Introduction	212
7.11.2. Significance of Region of Convergence in Stability	213

Contents

7.11.3. Poles and the Region of Convergence	213
7.11.4. Analytic Explanation	214
7.11.5. System function of Differentiator and Integrator	215
7.11.6. Further analysis of System Stability	215
7.11.7. Conclusion	216
7.12. Module 4: Lecture 12	
Stability in Laplace and Z-Domains	216
7.12.1. Introduction	216
7.12.2. Poles on the imaginary axis	216
7.12.3. Causality and Stability of Rational LSI system	217
7.12.4. Stability analysis in Z plane	217
7.12.5. Conclusion	219
7.13. Module 4: Lecture 13	
.	219
7.13.1. Introduction	219
7.13.2. Stability and Causality of Discrete System	219
7.13.3. Relationship between z-Transform and Laplace Transform	220
7.13.4. Formal Inverse of z-Transform	221
7.13.5. Conclusion	222
7.14. Module 4: Lecture 14	
Signal Flow Graphs	222
7.14.1. Introduction	222
7.14.2. Continuous Independent Variable Case	222
7.14.3. Discrete Independent Variable Case	225
7.14.4. General Form of Rational System Outputs	226
7.14.5. Conclusion	226

List of Figures

1.1.	4
1.2.	5
1.3. The δ_Δ function, centred at the origin	13
1.4. Representing the unit impulse graphically	14
1.5. Unit Impulse Response	14
1.6.	15
1.7. Consequence of shift-invariance of \mathbf{S}	15
1.8. Consequence of Linear shift-invariance of \mathbf{S}	15
1.9. RC circuit diagram	16
1.10.	17
1.11. $u(t)$ and $u(t + \Delta)$	18
1.12. Difference of $u(t)$ and $u(t + \Delta)$, divided by Δ	18
1.13. Shift-invariance	19
1.14. Additivity and Homogeneity	19
1.15. Scaling by Δ followed by taking the limit $\Delta \rightarrow 0$	19
1.16. Discrete Independent Variable System Representation	21
3.1. Applied input $X[n] = \delta[n]$	33
3.2. Applied delay to the signal	33
3.3. Scaling the input by $X[k]$	34
3.4. Summing the input over the entire region	34
4.1. Voltage generator (side view)	55
4.2. Sinusoidal Input To Stable LSI System	60
4.3. Plot of $x(t)$	68
4.4. Dot product of the square wave with cosines	68
4.5. Dot product of the square wave with sines	68
5.1. Sinc function	78
5.2. Symmetric Rectangular Pulse	80
5.3. Sinc Function	81
5.4. Fourier Transform of Rectangular Pulse at width T and $T/2$	81
6.1. Non-unique reconstruction due to lack of a-priori Information	103
6.2. True sinusoid and Ghost frequencies	106
6.3. The original signal, $y = \sin\left(\frac{2\pi t}{T_0} + \frac{\pi}{4}\right)$	108
6.4. The original signal, from Figure 6.3, sampled at points $t = nT_s, \forall n \in \mathbb{Z}$	109
6.5. Zoomed in view of the same signal	110
6.6. Missing 1 cycle and arriving at the correct edge	111
6.7. Missing 2 cycles and arriving at the correct edge	112
6.8. Missing 1 cycle and arriving at the incorrect edge	113
6.9. Missing 2 cycles and arriving at the incorrect edge	114

List of Figures

6.10. The red curve represents the sampled signal while the black curve represents the original signal	115
6.11. Block Diagram of a Sampling System.	119
6.12. Sampling of a Continuous Sine Wave	119
6.13. Principle of reconstruction	120
6.14. Desired frequency response of filter	121
6.15. Sampling at exactly twice f_m may yield all samples to be zero	121
6.16. Impulse response $h(t)$	122
6.17. Making use of the margin or the <i>don't care region</i> to construct a stable filter	123
6.18. Example of a filter which is not brickwall-like, but still serves our purpose	123
6.19. A filter which is practically realisable	125
6.20. RC circuit.	125
6.21. Magnitude response of RC circuit	126
6.22. The Pulse Train with period T_s	128
6.23. The 'sinc' function.	129
6.24. The sampled version of $\text{sinc}(r)$. r is on the x-axis. Some of the C_i 's are labeled.	129
6.25. $X(\Omega)$	130
6.26. Fourier transform when we disobey Nyquist principle: shows overlaps.	131
6.27. Fourier transform when we obey Nyquist principle: no overlap is seen.	131
6.28. A sampled function	134
6.29. Sample and Hold	134
6.30. Linear interpolation	135
6.31. Systematical representation of sample and hold	135
6.32. Impulse response of linear shift invariant system.	135
6.33. Fourier transform of $x(t)$	136
6.34. Spectrum of the sampled signal: $X_S(f)$	137
6.35. Hold frequency response	137
6.36. Superimposing hold freq. response on $X_s(f)$	137
6.37. Equalizer freq. response	139
6.38. Linear Interpolation of given points.	140
6.39. Passing $x(t)$ through the Sample and Hold Circuit.	140
6.40. Signal Representation along with the Sampling output and the Impulse response of LSI system	140
6.41. Signal Representation along with the Sampling output and the Impulse response of LSI system	141
6.42. Convoluting the Zero order hold with itself	141
6.43. Processing a Continuous Signal Using a DSP	144
6.44. Spectrum of a Speech Signal	144
6.45. Spectrum of the Sampled Speech Signal	144
6.46. A Digital System	145
6.47. Successive Approximation ADC	145
6.48. A Possible DAC	146
6.49. A Possible DAC	148
6.50. A Possible DAC	148
6.51.	149
6.52.	150
6.53.	151
6.54.	152

6.55. Region of relevance for $Y(\Omega)$ and $H(\Omega)$	154
6.56. Restriction on the frequency axis	155
6.57. Normalisation of the frequency axis with respect to sampling frequency . .	155
6.58. Digital equivalent of the continuous-time signal processing	156
6.59. Continuous-time processing that we wish to do	156
6.60. Frequency response of RC filter	157
6.61. Truncated frequency response to obtain the discrete-time frequency response	158
6.62. Normalised, restricted and periodically extrapolate the continuous-time frequency response to obtain the discrete-time frequency response	159
7.1. Region of convergence R_X for the input $x(t) = e^t u(t)$	169
7.2. Region of convergence R_X for the input $x(t) = e^t u(t)$	170
7.3. Region of convergence R_X for the input $x[n] = 3^n u[n]$	171
7.4. Region of convergence R_X for the input $x(t) = e^{2t} u(t) + e^{3t} u(t)$. (The green shaded region represents the ROC $\text{Re}(s) > 3$)	172
7.5. Region of convergence R_X for the input $x[n] = 2^n u[n] + 3^n u[n]$. (The green shaded region represents the ROC $ z > 3$)	173

List of Tables

Part I.

Natural Domain

1. Introduction to Abstractions

1.1. Module 1: Lectures 1 - 3

1.1.1. Introduction

With the advent of the internet, the amount of audio, video and image data has gone up. Processing this data for efficient storage and for a better human experience of them is a must. Apart from these types of data, there has also been a rise in data as describing the history of some phenomenon as a result of the numerous scientific experiments conducted and the world of financial markets. Data describing the temperature variation of a region over a region and data describing the variation of the price of a stock over time are examples. It is imperative to process this kind of data in order to infer from it. A very important characteristic of this kind of data is that it can be described as a function of a variable - discrete or continuous. All this data is just signals. And to process these varied types of signals, we construct systems. And it is these that we are going to hope to learn in this course. Below, we start with an introduction to abstraction, a process fundamental to the study of signals and systems.

1.1.2. Abstractions

The process of extracting out common attributes by doing away with the situation specific details irrelevant to the purpose at hand, from apparently different situations, is abstraction. The process of abstraction aids us by reducing our work as we can then conduct a study of all the situations all at once. It also provides us with an opportunity to apply insights from one specific situation to the others. All this is demonstrated by the example below.

We consider a mass m and a capacitor C . Let $F(t)$ be the force acting on the mass and let $v(t)$ be its velocity. From Newton's second law,

$$F(t) = m \frac{dv(t)}{dt}$$

In the case of the capacitor, if $i(t)$ is the current flowing through the capacitor and $v_C(t)$ is the voltage across it, then we know,

$$i(t) = C \frac{dv_C(t)}{dt}$$

We first notice that both of these are systems. The force $F(t)$ and the current $i(t)$ are the input signals to these systems and the velocity $v(t)$ and the voltage $v_C(t)$ are the corresponding output signals.

If one looks at the above relations carefully, one cannot help but notice the striking similarity between the two. If we consider the force $F(t)$, the mass m and the velocity $v(t)$ of the mass to be analogous to the current $i(t)$, the capacitor C and the voltage

1. Introduction to Abstractions

across the capacitor $v_C(t)$ respectively, the relations are the same. Now, we can just solve this one differential equation, forgetting conveniently the mass and the capacitor, and hence obtain the behaviour of both these systems at the same time. This is the power of abstraction.

Consider now, a viscous force proportional to the velocity of the mass with κ_0 being the constant of proportionality.¹ The force $F(t)$ now has to also overcome the opposing viscous force. Fig. 1(a) illustrates the situation. In the case of the mass, the equation describing the system now becomes,

$$F(t) = m \frac{dv(t)}{dt} + \kappa_0 v(t)$$

Also, consider a resistance R in series with the capacitor considered above. The circuit diagram is shown in fig. 1(b). For the RC circuit, the current flowing through the resistance is $C \frac{dv_C(t)}{dt}$, equal to the current flowing through the capacitor. By Kirchoff's law, $v_{in}(t)$ is equal to the sum of the voltages across the resistance and the capacitor. Thus,

$$v_{in}(t) = v_C(t) + RC \frac{dv_C(t)}{dt}$$

Dividing both sides by R ,

$$\frac{1}{R} v_{in}(t) = C \frac{dv_C(t)}{dt} + \frac{1}{R} v_C(t)$$

Again, there is a striking similarity between this relation and the relation for the mass system. Thus, it is clear that the two systems considered above are very similar in their behaviour.

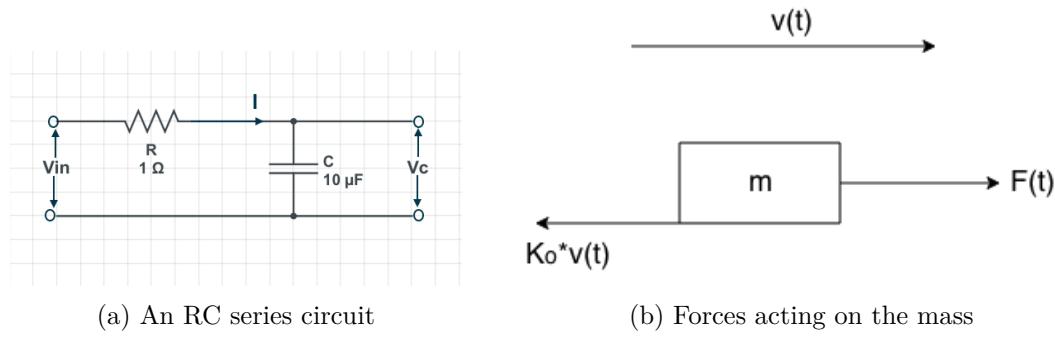


Figure 1.1.

1.1.3. Definitions

Signal

A signal is a “reasonable” function of a real independent variable to the set of complex numbers in the most general case. Thus a function $x(t)$ is a signal if $x : \mathbb{R} \rightarrow \mathbb{C}^2$ and $x(t)$ is reasonable. By reasonable, here we mean that the function should be well-behaved and

¹This is an approximation that is reasonably accurate at low velocities of the mass.

²The definition given here defines a signal with one dimensional independant and dependant variables.

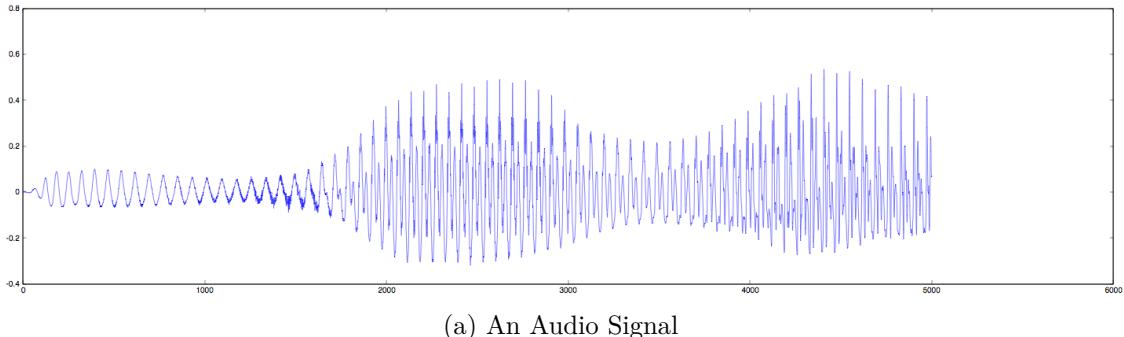
However signals with multidimensional independant variables are quite common as well. An image is an example with a two dimensional independant variable.

hence, often something we may encounter practically. For example, the Dirichlet function I_Q defined below is unreasonable in the sense that it is continuous nowhere on its domain and it is very improbable that this function will be encountered practically. Hence we do not consider this a signal.

$$I_Q(x) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational} \end{cases}$$

A signal is defined above to be a mapping to the set of complex numbers. The reason to consider a mapping to complex numbers is not yet clear as all the examples we have considered were real valued. However, we note that the definition given is a more general one as the set of reals is a subset of the set complex numbers. The reason for this generality will become clear as we go further in the course.

A couple of examples of signals are given in fig. 2 below.



(a) An Audio Signal



(b) An image: A 2D Signal

Figure 1.2.

System

A system is a mapping from a set of signals to another set of signals. A system thus takes a signal as an input and returns another signal as its output. A microphone is an

1. Introduction to Abstractions

example of a system. This system is a mapping from the set of acoustic signals to a set of electric signals.

System Description

Given a system, a system description is an explicit or an implicit relation between the inputs and the outputs of the system. By an implicit relation, we mean a relation of the form $F(x(t), y(t)) = 0$ where F is a function of two signals, the input signal $x(t)$ and the output signal $y(t)$. Thus it is not always possible to directly compute $y(t)$ directly in terms of $x(t)$. Description 1 given below is an explicit system description while description 2 is an implicit one.

$$1 \quad y(t) = x(t) + 3\frac{dx(t)}{dt}$$

$$2 \quad y(t)x(t) + x(t)\frac{dy(t)}{dt} = 0$$

1.1.4. Properties of Systems

Given a system description, a lot can be inferred about the system. We describe below some properties that systems may or may not possess and give examples of each. These properties aid in the further study and classification of systems.

Additivity

A system S is said to be additive if for any two input signals $x_1(t)$ and $x_2(t)$,

$$x_1(t) \xrightarrow{S} y_1(t) \text{ and } x_2(t) \xrightarrow{S} y_2(t)$$

imply

$$x_1(t) + x_2(t) \xrightarrow{S} y_1(t) + y_2(t)$$

The systems $y(t) = 3x(t)$ and $\frac{dy(t)}{dt} = x(t)$ are both additive while the system $y(t) = x(t) + 4$ (a DC shift in the case of electrical signals) is not because of the constant.

Homogeneity

A system S is said to be homogenous if for every input signal $x(t)$ and for every $c \in \mathbb{C}$,

$$x(t) \xrightarrow{S} y(t)$$

implies

$$cx(t) \xrightarrow{S} cy(t)$$

This property is also referred to as the scaling property.

The systems $y(t) = 3x(t)$ and $\frac{dy(t)}{dt} = x(t)$ are both homogeneous too while the system $y(t) = x(t) + 4$ is non-homogeneous again due to the presence of the non zero constant.

Shift Invariance

A system S is said to be shift invariant if for all input signals $x(t)$ and for every $t_0 \in \mathbb{R}$,

$$x(t) \xrightarrow{S} y(t)$$

implies

$$x(t - t_0) \xrightarrow{S} y(t - t_0)$$

Intuitively, the property of shift invariance implies that if the same input is fed to the system at different times, then the output of the system will only be shifted in time by the same amount as that of the input.

Thus, the system $y(t) = x(t^2)$ is shift invariant while the system $y(t) = tx(t)$ is not.

Note: The property of shift invariance is also referred to by many as time invariance.

Some Additional Material

Abstraction

As the professor alludes to in the video segment, the process of abstraction is the most important tool we use for analysing and studying the natural world. A very good example of abstraction is the theory of gravitation. The insight that an apple falling from the tree (and supposedly hitting his head) and the Moon revolving around the Earth are actually the same thing was something only Newton could have had. In his words, “....I began to think of gravity extending to the orb of the Moon....”. At that moment in history, he abstracted out all the differences between the moon and the apple to come up with the theory of universal gravitation. The word universal in the name is important as it states how we can think of (i.e abstract) every mass in the Universe as a point mass attracting every other. That is the power of abstraction as the instructor stresses in the video segment.

However one must keep in mind that the real art lies in finding the right kind and the right level of abstraction for the specific kind of job at hand. Suppose a chemist wants to study chemical behaviour of ethanol and methanol in connection with their reactions with some particular chemical. In that case, abstracting out the fact whether it is methyl or ethyl alcohol and studying, in generality the reactions of the alcohol part of the two (i.e. of the -OH group) with the desired chemical considerably reduces the work of the chemist. However, we cannot use the same abstraction when studying the biological effects of the two as methanol is highly toxic and can be fatal in sufficient quantities; however ethanol is not toxic and is used for....well we all know!

1.2. Module 1: Lecture 4

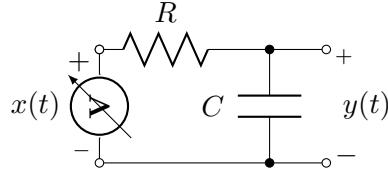
System Properties: An Illustration

1.2.1. Introduction

Recall the definitions of additivity, homogeneity and shift-invariance. In the last lecture, we saw that a system with the description, $y(t) = x(t) + 5$ is neither additive nor homogeneous. Is it shift-invariant? It is because applying the definition of shift-invariance, it can be seen that $y(t - t_0) = x(t - t_0) + 5$. In this lecture, we will consider a real-life example of a system and apply the definitions to analyse its additivity, homogeneity and shift-invariance.

1. Introduction to Abstractions

1.2.2. RC Circuit



Let us derive the system description for this RC circuit. Applying Kirchoff's Voltage law, we get $x(t) - i(t)R - y(t) = 0$ where $i(t) = C \frac{dy(t)}{dt}$. So we get,

$$x(t) = RC \frac{dy(t)}{dt} + y(t)$$

There is a way to check whether both additivity and homogeneity are simultaneously satisfied, using the test of *superposition*. A system is said to obey the principle of superposition if any *linear combination of the inputs gives the same linear combination of the outputs*. Consider two different inputs $x_1(t)$ and $x_2(t)$, giving outputs $y_1(t)$ and $y_2(t)$ respectively. We ask whether, $\alpha x_1(t) + \beta x_2(t)$ yields $\alpha y_1(t) + \beta y_2(t)$ **for all** possible x_1, x_2, α and β ? If so, the system obeys superposition and hence is both additive and homogeneous. Let us apply it to the RC circuit system.

$$x_1(t) = RC \frac{dy_1(t)}{dt} + y_1(t) \quad (1.1)$$

$$x_2(t) = RC \frac{dy_2(t)}{dt} + y_2(t) \quad (1.2)$$

Multiplying (1.1) by α and (1.2) by β and adding, we get

$$\begin{aligned} \alpha x_1(t) + \beta x_2(t) &= RC \frac{d\alpha y_1(t)}{dt} + \alpha y_1(t) + RC \frac{d\beta y_2(t)}{dt} + \beta y_2(t) \\ &= RC \frac{d(\alpha y_1(t) + \beta y_2(t))}{dt} + (\alpha y_1(t) + \beta y_2(t)) \end{aligned} \quad (1.3)$$

Since x_1, x_2, α and β are arbitrary, it holds for all possible values of these. Therefore, the RC system obeys principle of superposition.

The physical interpretation of superposition is that inputs are scaled and then added one on top of the other, or in other words *superposed*. If the outputs also undergo the same scaling and addition, **for all** possible scaling factors and inputs, the system obeys the principle of superposition. The principle of superposition subsumes additivity and homogeneity. That is, if the principle of superposition holds, additivity and homogeneity will both hold.

Additivity

In the principle of superposition, substitute α and β both equal to 1. We get back the property of additivity. Does $x_1(t) + x_2(t)$ yield $y_1(t) + y_2(t)$ for all possible x_1 and x_2 ?

Homogeneity

In the principle of superposition, substitute $\beta = 0$. We get back the property of homogeneity. Does $\alpha x_1(t)$ yield $\alpha y_1(t)$ for all possible x_1 and α ?

Now that it is proved that the system is superposable, let us ask the question - *What changes can we make to the properties of the RC circuit system to destroy its superposability?*

1.2.3. Introduction

Recall the definitions of additivity, homogeneity and shift-invariance. In the last lecture, we saw that the RC circuit system obeys the principle of superposition. That is, it is both additive and homogeneous. In this lecture, we will see how this can be destroyed by small changes and also analyse its shift-invariance property.

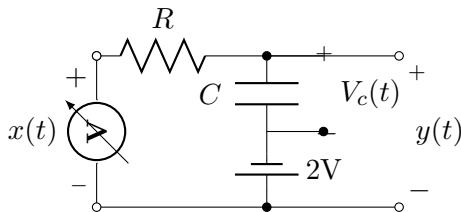
1.2.4. RC Circuit

We first start with the question : *What changes can we make to the properties of the RC circuit system to destroy its superposability?* One possibility is to introduce a little nonlinearity to the system by making the response of the resistor different.

For example, instead of $V \propto i$, we could have $V \propto i^\gamma$. This is quite true since real resistances do have a nonlinear regime. In microelectronics, resistances are made of semiconductor devices which do not have the ideal Ohm's law behaviour. Let us consider what would happen to the system in this case. Using the Kirchoff's voltage law like last time, the system description can be derived to be

$$x(t) = R(C \frac{dy(t)}{dt})^\gamma + y(t)$$

It is left as an exercise for the reader to show that this system no longer obeys superposition. This implies that the system is not *both* additive and homogeneous. One must also show that the system is *neither* additive nor homogeneous. This question has been raised to point to the fact that all real systems behave linearly over a regime and not always.



Another way in which superposition can be destroyed is by adding a *DC offset*. This can be introduced by adding a DC voltage source in series with the capacitor. Earlier, we had the capacitor voltage $V_c(t) = y(t)$. Now, we have $V_c(t) + 2 = y(t)$. The system description can be written using Kirchoff's voltage law. $x(t) - i(t)R - V_c(t) - 2 = 0$, where $i(t) = C \frac{dV_c(t)}{dt}$. So we get,

$$x(t) = RC \frac{dV_c(t)}{dt} + V_c(t) + 2 \quad (1.4)$$

$$x(t) = RC \frac{dy(t)}{dt} + y(t) \quad (1.5)$$

This is the same as before and so it obeys principle of superposition. Now, if we were to consider the output to be $V_c(t)$ and not $y(t)$, the system description would be (1.4). This is similar to the system description $y(t) = x(t) + 5$. Using the same technique as before, the reader has to show that the system described by (1.4) is neither additive nor homogeneous.

1. Introduction to Abstractions

Shift-invariance

Is the system described by (1.5) is shift-invariant? In other words, is it true that

$$x(t - t_0) = RC \frac{dy(t - t_0)}{dt} + y(t - t_0)?$$

Yes. This is because the derivative operator is shift-invariant since $\frac{dy(t-t_0)}{dt} \equiv \frac{dy(\lambda)}{d\lambda}$, where $\lambda = t - t_0$.

Question: Is the system described by (1.4) shift-invariant?

1.3. Module 1: Lecture 5

Challenging Problems

1.3.1. Introduction

So far, we have looked at systems which are either both additive and homogeneous or neither additive nor homogeneous. But they are independent properties. In this lecture, we shall look at examples of *complex-valued* systems which are additive but not homogenous and vice versa.

1.3.2. Two systems

Complex-valued systems are those whose inputs and outputs are complex functions of time. That is, $x(t)$ and $y(t)$ take values belonging to \mathbb{C} (the set of complex numbers).

System 1 :

$$y(t) = \operatorname{Re}(x(t))$$

Is this system additive? Is it homogeneous? (Remember that the constant α can be complex)? Is it shift-invariant?

System 2 :

$$y(t) = \begin{cases} \frac{x(t)x(t-1)}{x(t-2)}, & \text{if } x(t-2) \neq 0 \\ 0, & \text{otherwise} \end{cases}$$

Is this system additive? Is it homogeneous? Is it shift-invariant?

The systems are, in fact, shift-invariant as shifting the input by any constant t_0 in time, will result the output to also shift by the same amount in time. What can we do to the system to make them *shift-variant*? Note that shift-invariance can be destroyed by having a time-dependency in the system description. Here is an example of explicit time-dependence.

$$y(t) = tx(t)$$

When the input is shifted by t_0 , we get the output to be $tx(t - t_0)$ which is not equal to $y(t - t_0) = (t - t_0)x(t - t_0)$.

Systems with the three properties of additivity, homogeneity and shift-invariance are very important because (a) they are easy to analyse and (b) they are easy to realise. Any electrical system consisting of ideal resistors, capacitors and inductors obeys these

1.4. Module 1: Lecture 6 Impulse Response: A Building Block for Continuous Functions

three properties. Mechanical systems consisting of springs and masses obey these three properties. For example, a hydraulic system can be modeled as the RC circuit that we previously saw. A tank which stores water is analogous to a capacitor which stores charge; a pipe which allows water to flow between its two ends is analogous to a resistor which allows current to pass through it.

Challenge : Build a system description of a tank analogous to the capacitor using basic fluid equations. *Hint* - Let the pressure difference between the top and bottom of the tank $P_0(t)$ be the output and the height of water in the tank $h(t)$ be the input. The derivative relation arises from the speed of water $v_1(t)$ and the instantaneous height $h(t)$ as $v_1(t) = \frac{dh(t)}{dt}$.

Challenge : Build a system description of a pipe analogous to the resistor using basic fluid equations. *Hint* - Let the pressure difference across the pipe $P_a(t) - P_b(t)$ be the output and the velocity of the fluid through the pipe $v_2(t)$ be the input. There is a linear relationship between the two.

Challenge : By attaching the pipe to the bottom of the tank, build a system description of the tank and pipe system analogous to the resistor-capacitor system.

Hint - Let the input be the difference in the pressures between the top of the tank to the free end of the pipe. Let the output be the pressure difference across the tank.

1.4. Module 1: Lecture 6

Impulse Response: A Building Block for Continuous Functions

1.4.1. Introduction

As mentioned in the previous lecture, systems which obey the three properties of additivity, homogeneity and shift-invariance are desirable since they are easy to analyse and build. In this lecture, we will make a start to understand why this is so, by introducing the concept of narrow pulses.

1.4.2. Narrow pulses

Narrow pulses can be used to construct *reasonable signals*. Reasonable signals or smooth signals are those which do not have infinite number of discontinuities in a finite interval. Therefore, there exists an interval where the signal is continuous. Consider the signal over such an interval.

Divide the interval into sub-intervals of size as small as can be. Consider one such sub-interval. As the width of this sub-interval grows smaller and smaller, the value that the signal takes over it does not change much and is almost a constant.

Consider a narrow pulse of width Δ . Let its height be $\frac{1}{\Delta}$, such that its area is unity. This is our basic building block. When this pulse is placed at a point in time t and multiplied by the signal, we get a pulse of height *proportional* to the (constant) value of the signal in that interval (the proportionality factor being $\frac{1}{\Delta}$).

Now copy and shift the initial pulse (of width Δ and height $\frac{1}{\Delta}$) in the time-axis in steps of Δ so as to cover the entire time-axis. Multiply it by the signal and combine all of it

1. Introduction to Abstractions

together. We get back something very similar to the original signal. Smaller the pulse width, more the similarity to the original signal.

1.4.3. Sifting property

Let $x(t)$ be the continuous function. Reconsider the narrow pulse that we constructed in the previous lecture, placed at $t = 0$. It has width Δ and height $\frac{1}{\Delta}$. This can be called as $\delta_\Delta(t)$.

Place one such pulse (red) at $t = t_0$ on the time-axis as shown in figure ..It is $\delta_\Delta(t - t_0)$. Multiply $x(t)$ with $\delta_\Delta(t - t_0)$ and integrate over all time t . That is,

$$\int_{-\infty}^{\infty} x(t)\delta_\Delta(t - t_0)dt \quad (1.6)$$

If Δ is small enough, the function over the non-zero part of the pulse is almost a constant which can be taken to be $x(t_0)$. So the above integration simply picks out the value of the function at one point t_0 . This can be seen in the following.

$$x(t)\delta_\Delta(t - t_0) \equiv x(t_0)\delta_\Delta(t - t_0) \quad (1.7)$$

This is true because the product of the function and the narrow pulse is zero everywhere except over the small interval around t_0 , where the function is almost a constant. Now keeping in mind that the area of $\delta_\Delta(t)$ pulse is unity, if we were to perform the integration as (1.6), we would get,

$$\begin{aligned} \int_{-\infty}^{\infty} x(t)\delta_\Delta(t - t_0)dt &= \int_{-\infty}^{\infty} x(t_0)\delta_\Delta(t - t_0)dt \\ &= x(t_0) \int_{-\infty}^{\infty} \delta_\Delta(t - t_0)dt \\ &= x(t_0) \end{aligned} \quad (1.8)$$

So the $\delta_\Delta(t - t_0)$ pulse is like a sieve which sifts or picks out the value of the function at t_0 .

1.4.4. Stitching together a function with narrow pulses

We can derive a different interpretation of this integration property. Construct the $\delta_\Delta(t)$ pulse to be symmetric about the t -axis such that $\delta_\Delta(-t) = \delta_\Delta(t)$. Since,

$$\int_{-\infty}^{\infty} x(t)\delta_\Delta(t - t_0)dt = x(t_0) \quad (1.9)$$

it is true that,

$$\int_{-\infty}^{\infty} x(t)\delta_\Delta(t_0 - t)dt = x(t_0) \quad (1.10)$$

The meanings of the equations (1.9) and (1.10) are very different. (1.9) says that multiplying a function by the narrow pulse and integrating over all time results in pulling out the value of the function at the point at which the pulse is located.

For interpreting the meaning of (1.10), consider t_0 to be the basic variable by which the function is indexed/described. Take any point t and multiply the value of the function

at that point $x(t)$ with the narrow pulse at that point. Then take a sum over all such t . This sum tends to an integral in the limit. This integral which is the limit of this sum is precisely given by (1.10). It can be thought of as,

$$x(t_1)\delta_\Delta(t_0 - t_1) + x(t_2)\delta_\Delta(t_0 - t_2) + x(t_3)\delta_\Delta(t_0 - t_3) + \dots \quad (1.11)$$

Graphically, it means, Put a narrow pulse at every possible t_0 and multiply with the function value at that point. So each pulse will have a height = (value of the function at that point) ($\frac{1}{\Delta}$). As there is a continuum of pulses, combining the pulses leads to the integral whose value is nothing but the function itself, in the limit $\Delta \rightarrow 0$. This gives the interpretation that **Every function is a combination of an infinite number of very narrow pulses.**

Note that these narrow pulses are such that even as Δ goes to zero, the height grows as $\frac{1}{\Delta}$, keeping the area always equal to unity.

1.5. Lecture 7. Impulse Response and Properties of Systems

1.5.1. The Continuous Variable Delta Function

We defined the continuous time delta function as a function which assumes the value 0 everywhere except in an interval Δ around the origin, where it assumes the value $\frac{1}{\Delta}$, and then we take the limit as $\Delta \rightarrow 0$. Refer to fig. 1.

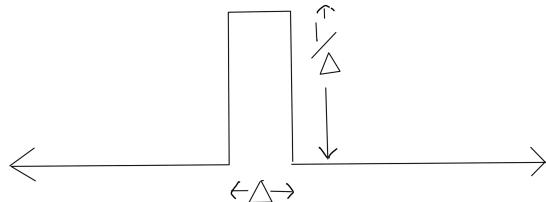


Figure 1.3.: The δ_Δ function, centred at the origin

Thus, the area under the function, i.e. the integral of the function from $-\infty$ to $+\infty$ is unity, while the function is zero at all points except the origin.

It should be noted that the delta function is not a function in the conventional sense. This is because, the delta function does not have a well-defined value at the origin. We call the delta function a "generalized function". Generalized functions are generalizations or extensions of the concept of a function, but are not really functions.

We will also refer to the delta function as a "unit impulse". The unit impulse captures an area of unity while having zero width. The conventional way of representing a unit impulse is shown in fig. 2.

1. Introduction to Abstractions

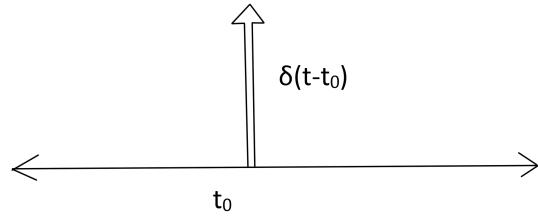


Figure 1.4.: Representing the unit impulse graphically

Now, if we use the symbol $\delta_\Delta(t)$ for the above mentioned function with finite Δ , then in accordance with the definition above,

$$\lim_{\Delta \rightarrow 0} \delta_\Delta(t) = \delta(t)$$

Where, we denote by $\delta(t)$ the unit impulse function.

1.5.2. Unit Impulse Response

Assume that we give to a system, a unit impulse/delta function as its input. The output that is obtained is called the impulse response of that system.

Although practically, we cannot generate an impulse function, one can imagine giving to the system δ_Δ functions as inputs, successively, with decreasing values of Δ and obtain a good estimate of the Unit Impulse Response.

Thus, in short, **the impulse response is the output of a system, when the input is a unit impulse.** (Fig. 3)

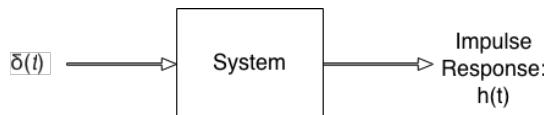


Figure 1.5.: Unit Impulse Response

Impulse response of an LSI system

Consider a linear shift-invariant system S . Now, suppose that we give as the input, the sum of two unit impulses centred at two different points t_1 and t_2 . Due to additivity, we expect the output to be the sums of the corresponding outputs. (fig. 4(a))

Also, because of homogeneity, we expect that the output corresponding to the input $\kappa\delta(t)$ should be $\kappa h(t)$ where $h(t)$ is the unit impulse response and $\kappa \in \mathbb{C}$ (fig. 4(b))

1.5. Lecture 7. Impulse Response and Properties of Systems

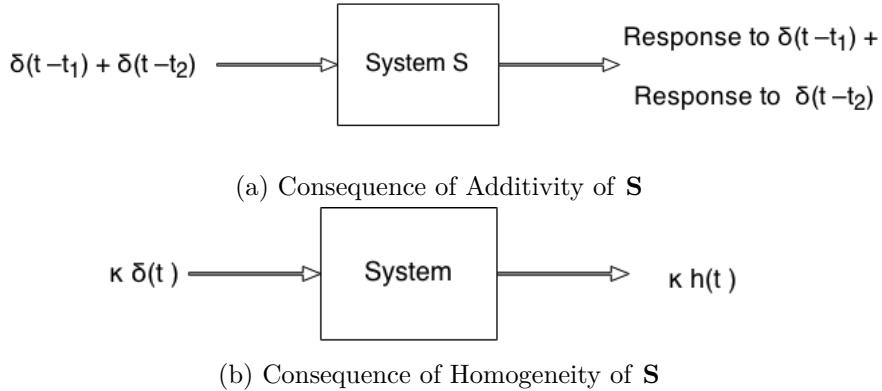


Figure 1.6.

Now, if we took κ to be the value of a function x at some point, then the property still holds.

Let us not consider the output of $\delta(t-t_0)$, which is nothing but $\delta(t)$ shifted to the right by t_0 units. Thus, due to shift invariance, we expect that the output to $\delta(t-t_0)$ should be $h(t-t_0)$. (fig. 5)

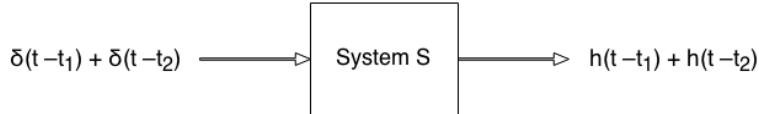


Figure 1.7.: Consequence of shift-invariance of **S**

Now assume that we have points λ_1 , λ_2 and λ_3 . As a consequence of Homogeneity, Additivity and Shift-Invariance of **S**, we know that the output corresponding to $x(\lambda_1)\delta(t-\lambda_1) + x(\lambda_2)\delta(t-\lambda_2) + x(\lambda_3)\delta(t-\lambda_3)$ will be $x(\lambda_1)h(t-\lambda_1) + x(\lambda_2)h(t-\lambda_2) + x(\lambda_3)h(t-\lambda_3)$.

We can think of an integral as a summation of many such products $x(\lambda)\delta(t-\lambda)$ where the λ 's are brought infinitesimally close to each other. Thus, we can consider the integration as a summation over a continuum of λ 's. (fig. 6)

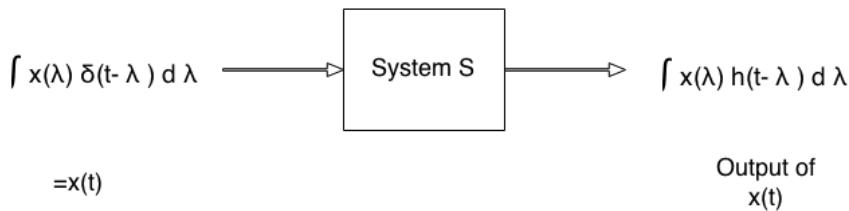


Figure 1.8.: Consequence of Linear shift-invariance of **S**

As discussed in the previous lecture, the input integral equates to $x(t)$. Thus, we have obtained a way of obtaining the output corresponding to any signal $x(t)$ using the impulse response.

Given the impulse response of an Linear Shift-Invariant System, we can calculate the output corresponding to any input signal.

1. Introduction to Abstractions

This is a very important result in signals and systems, because knowing the output of just one function gives us the power to know beforehand, what the output should be for any input provided to the LSI system.

1.6. Lecture 8: Relating Unit Step and Unit Impulse

1.6.1. Impulse Response of for an RC circuit

In the discussion about the importance of abstraction, we had come across two systems, the RC circuit, and the mass-force system, which had the same behaviour. Now, since the RC circuit system is a Linear Shift-Invariant system, it would be useful to acquire its unit impulse response, since this will empower us to obtain the output signal, for any given input signal.

Recall that for the RC circuit system, we took the voltage applied across the resistor and the capacitor to be the input voltage (V_{in} - the input signal), while the voltage across the capacitor to be the output voltage (V_C - the output signal). (fig. 1)

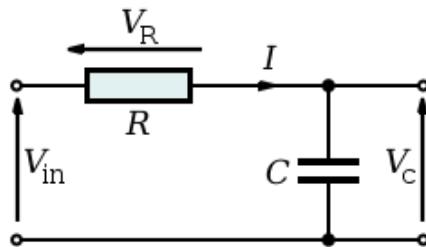


Figure 1.9.: RC circuit diagram

Now, we had defined the unit impulse as a limiting case of the δ_Δ function, whose width was Δ , and whose area under the curve was unity.

To find the unit impulse response, we need to go back to the differential equation governing the RC circuit. We know that the current in the circuit is proportional to the derivative of the capacitor voltage. i.e.

$$I = C \frac{dV_C}{dt}$$

On integrating both the sides w.r.t. t, we get

$$\int Idt = CV_C$$

Instead of dealing with the Voltage directly, we will look at the integrated form. Thus arises the question: What is the integral of the unit impulse?

Integral of the unit impulse

Consider the δ_Δ function centred at t_0 . Now consider the running integral:

$$\text{Running Integral of } \delta_{\Delta,t_0} = \int_{-\infty}^t \delta_{\Delta,t_0}(\lambda) d\lambda$$

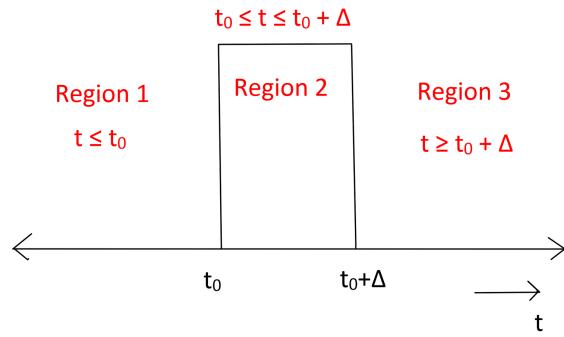
Let us now divide the δ_{Δ,t_0} function into three regions (fig. 2(a)):

1.6. Lecture 8: Relating Unit Step and Unit Impulse

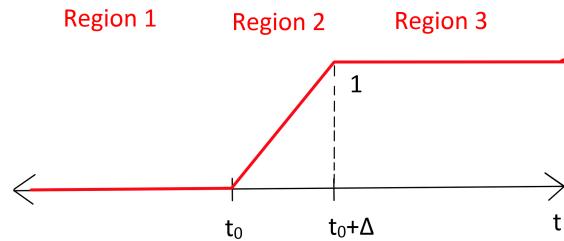
- Region 1: The region $t < t_0$
- Region 2: The region where $t_0 < t < t_0 + \Delta$
- Region 3: The region $t > t_0 + \Delta$

We can see that the integral in Region 1 is zero since δ_{Δ, t_0} is 0 and that the only region where we need to calculate the integral is Region 2.

The integral of a constant function is a linear function, and since we know that the area under the δ_Δ function is unity, it is clear that the running integral increases linearly from 0 at $t = t_0$ to 1 at $t = t_0 + \Delta$ after which, it remains 1 in Region 3. (fig. 2(b))



(a) Dividing the δ_Δ function into 3 regions



(b) Running integral of delta function

Figure 1.10.

Now, to find the unit impulse response, we need to let $\Delta \rightarrow 0$. Clearly, Region 2 vanishes and we get the unit step, located at t_0 . Thus,

The running integral of the unit impulse is the unit step function, and the derivative of the unit step function is the unit impulse³

$$\text{Unit Impulse} \xrightarrow{\text{running integral}} \text{Unit Step}$$

$$\text{Unit Impulse} \xleftarrow{\text{derivative}} \text{Unit Step}$$

Now, the derivative of a function $x(t)$ is defined as:

$$\frac{dx(t)}{dt} = \lim_{\Delta \rightarrow 0} \frac{x(t + \Delta) - x(t)}{\Delta}$$

³Note that in the classical sense of the derivative, it is not possible to take the derivative of the unit step at the origin, since it is a discontinuous function. Just as the unit impulse is not really a function, but a generalized function, the derivative of the unit step function is not a usual derivative, but is a ‘generalized derivative’. However, without going in the subtleties of the definition of a generalized derivative, we shall move ahead without distinguishing between the classical and generalized derivatives

1. Introduction to Abstractions

The term $x(t + \Delta)$ can be considered as the function $x(t)$ shifted to the left by Δ units. Thus, the derivative can be thought of as the difference of the original function and a shifted version of the same function, divided by the shift in the limit of the shift going to zero. This is a more ‘Signals and Systems’ way of thinking of derivatives of signals. The original aim of this exercise was to find the unit impulse response of the RC circuit. But instead of finding the unit impulse response, let us first find the unit step response. It can be shown that the expression for the unit step response will be

$$(1 - e^{-\frac{t}{\tau}})u(t)$$

where, $\tau = RC$ and $u(t)$ is the unit step function. This can be derived using knowledge of differential equations and can be verified by substituting in the RC circuit equation. Physically, we can argue that the potential across the capacitor (i.e. the output voltage) cannot rise suddenly, and thus increases gradually and asymptotically towards a maximum value, for $t > 0$.

Now let us come back to the derivative of the unit step function. If we denote the unit step function by $u(t)$, its derivative will be:

$$\frac{du(t)}{dt} = \lim_{\Delta \rightarrow 0} \frac{u(t + \Delta) - u(t)}{\Delta}$$

This is the difference between $u(t)$ and its shifted version, divided by Δ (fig. 3(a))

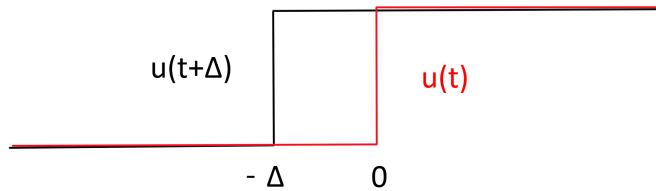


Figure 1.11.: $u(t)$ and $u(t + \Delta)$

Now, taking their difference and dividing by Δ (fig. 3(b)), we get:

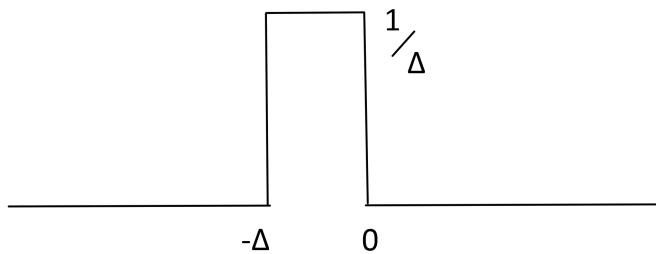


Figure 1.12.: Difference of $u(t)$ and $u(t + \Delta)$, divided by Δ

which is nothing but the δ_Δ function which we talked about. Taking the limit $\Delta \rightarrow 0$, we get the derivative of $u(t)$ on one side, and the unit impulse $\delta(t)$ on the other. Thus, there is a relationship of integration and differentiation between the unit step function and the unit impulse function. We have also found out the unit step response of the RC circuit. Since the derivative of the unit step function is the unit impulse function, is it true that the derivative of the unit step response for the RC circuit will give the unit

impulse response? More generally, if an input $x(t)$ produces an output $y(t)$ for an LSI system, does the derivative of $x(t)$ produce as the output, the derivative of $y(t)$?

Intuitively, it seems to be correct, and taking the derivative of the unit step response for the RC circuit should give its unit impulse response. This question will be answered in the next lecture.

1.7. Lecture 9: Unit Impulse Response of the RC circuit

1.7.1. Taking the derivative of the input signal for an LSI system

Let \mathbf{S} be a Linear Shift-Invariant system, and let $y(t)$ be the output corresponding to an input signal $x(t)$.

Because of the Shift invariance of the \mathbf{S} , $x(t + \Delta)$ will produce an output of $y(t + \Delta)$

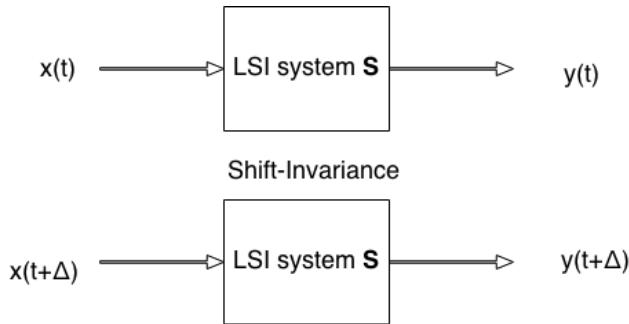


Figure 1.13.: Shift-invariance

Homogeneity implies that $-x(t)$ will result in an output of $-y(t)$. And thus using additivity, an input of $x(t + \Delta) - x(t)$ will produce an output of $y(t + \Delta) - y(t)$.

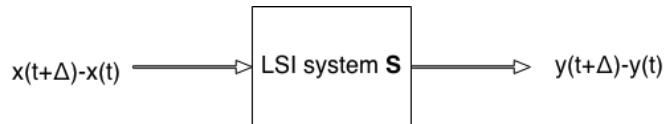


Figure 1.14.: Additivity and Homogeneity

Invoking homogeneity with a scaling of $\frac{1}{\Delta}$, and taking the limit $\Delta \rightarrow 0$, we obtain the result that the derivative of $x(t)$ produces the derivative of $y(t)$, as the output.

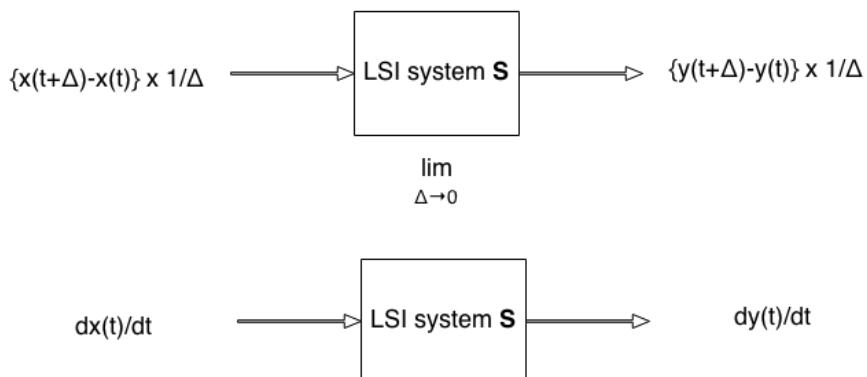


Figure 1.15.: Scaling by Δ followed by taking the limit $\Delta \rightarrow 0$

1. Introduction to Abstractions

Impulse response of the RC circuit

Now we can finally obtain the impulse response of the RC circuit. To do this, we need to apply the result of the previous section to the the unit step response of the RC circuit. As discussed previously, the unit step response of the RC circuit is the function $s(t) = (1 - e^{-\frac{t}{\tau}})u(t)$.

Now, since the RC circuit is an LSI system, the derivative of $u(t)$ should produce the derivative of $s(t)$! But we already know that the derivative of the unit step function is the unit impulse. Thus the unit impulse response is nothing but the derivative of the unit step response $s(t)$!

Using the product rule for differentiation, we get

$$\begin{aligned}\frac{ds(t)}{dt} &= u(t) \left\{ \frac{d}{dt} \left(1 - e^{-\frac{t}{\tau}} \right) \right\} + \left(1 - e^{-\frac{t}{\tau}} \right) \frac{du(t)}{dt} \\ &= \frac{1}{\tau} e^{-\frac{t}{\tau}} u(t) + \left(1 - e^{-\frac{t}{\tau}} \right) \delta(t)\end{aligned}$$

Now note that the second term in the expression is the delta function multiplied by another function. But $f(t)\delta(t) = f(0)\delta(t)$. The second term in the above equation vanishes, since $f(0)=0$ in this case. Thus we obtain,

$$\frac{ds(t)}{dt} = \frac{1}{\tau} e^{-\frac{t}{\tau}} u(t)$$

This is the formula for the Impulse Response of an RC circuit.

Interpretation of the RC unit Impulse Response

We can try to understand how this result makes sense physically.

Applying a unit impulse voltage across the RC voltage is like suddenly pumping in an amount of charge into the circuit. Let us use the δ_Δ function instead of the the unit impulse, since it is easier to interpret applying a finite voltage $\frac{1}{\Delta}$ over a finite time Δ . Since the voltage across a capacitor does not behave in an ‘impulsive’ manner, this sudden jump in the voltage at $t = 0$ is sustained by the resistor R, setting up a current:

$$I = \frac{1}{\Delta} \cdot \frac{1}{R}$$

Now, at the same time, the same current flows through the Capacitor, and integrating this current gives us the capacitor charge:

$$\text{Capacitor Charge} = \Delta \cdot \frac{1}{\Delta} \cdot \frac{1}{R}$$

Resulting in a voltage across the Capacitor (at time $t = 0$):

$$\begin{aligned}\text{Capacitor Voltage at time } t=0 &= \Delta \cdot \frac{1}{\Delta} \cdot \frac{1}{R} \cdot \frac{1}{C} \\ &= \frac{1}{RC} \\ &= \frac{1}{\tau}\end{aligned}$$

Since there is no applied voltage across the circuit after $t = 0$, the capacitor simply discharges exponentially, thus giving us :

$$\text{Capacitor voltage, i.e. output voltage} = \frac{1}{\tau} e^{-\frac{t}{\tau}} u(t)$$

1.8. Lecture 10: Introduction to Discrete Systems

1.8.1. Introduction

In the previous sessions, we have learnt about the continuous time systems. We also studied the different properties of these systems like additivity, homogeneity, shift invariance and so on. We are no able to identify whether a given continuous time system is additive or shift invariant or homogenous or not. However, all these parameters were studied only with respect to a certain type of systems i.e, Continuous Time Systems. These are the systems where the independent variable is continuous and the signal is defined for a continuous range of values. We will now look at a different type of systems i.e, the Discrete 'Independent Variable' Systems. It should be noted that a very important class if discrete signal arises from the sampling of continuous time signal.

1.8.2. Definition

Systems where in the independent variable assumes only discrete values are known as discrete 'independent variable' systems.

Examples A very well known example of this kind of system is the Stock Market Index. Another example could be that of a graph showing the scoring and run rate of a batting team in a cricket match.

1.8.3. Representation

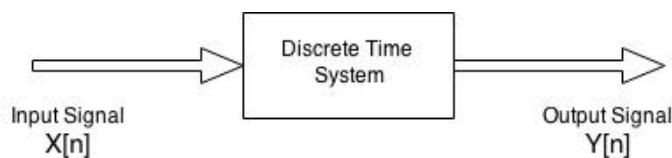
Consider a co-ordinate system of only one axis. Now the discrete time system independent variable will only assume discrete values. Thus, the axis will have the co-ordinates separated by nT , where n is an integer and T is the 'sampling time period' and can be one second, one day, one year and so on.

Once we decide ' T ', only it's value matters.

We shall represent a Discrete System as:

$$X[n] \rightarrow Y[n]$$

We shall use the square brackets [] to show discreetness of the system.



Σ

Figure 1.16.: Discrete Independent Variable System Representation.

1. Introduction to Abstractions

1.8.4. Examples of a Discrete Time System

Example 1

Consider a Bank Account.

A certain amount $X[n]$ is deposited at a given time.

T is the interval of interest calculation.

$Y[n]$ is the amount present at the n th instant.

The Bank gives an interest half yearly. Thus, a certain amount of interest is calculated after 6 months. The next amount is calculated over the next 6 months.

α is the rate of interest for the first 6 months while β will be the rate of interest for the next 6 months.

Thus if α and β are the interest rates over two intervals of 6 months each.

The amount $Y[n]$ at the n th instant will be given as:

$$Y[n] = (\alpha \cdot Y[n - 1]) + (\beta \cdot Y[n - 2]) + X[n]$$

The amount $Y[n]$ will thus be the interest collected over the previous 2 intervals of 6 months each along with the current amount in the account $X[n]$.

This can be seen as an example of a Discrete System.

Also, since the output is dependent on it's own past, it is a 'Recursive System'.

Example 2

Consider a Population in a region.

We need to calculate the Tax Paid $Y[n]$ by this population.

Population (Input) is given by $X[n]$.

The tax is paid over two intervals at the rates α and β .

The above system can be stated as:

$$Y[n] = (\alpha \cdot X[n]) + (\beta \cdot X[n - 1])$$

The above system output or The Tax collected is Independent of it's previous state.

Hence above is a Non-Recursive System.

The Properties of Discrete Independent Variable systems will be studied in the following sessions.

1.9. Lecture 11: Properties of Discrete Systems.

1.9.1. Additivity

Introduction

Previously, we were introduced to a new form of systems that is, Discrete Independent Variable Systems. If we recall, we had established a fact that the properties of Discrete Systems are more or less similar to those of the Continuous Systems. In this session, we look at one of the properties of Discrete Systems i.e, Additivity.

Definition

Additivity can be defined such that a system is said to be additive if the output corresponding to the sum of any two inputs is the sum of the two outputs.

Thus if we have two inputs to a given system $X_1[n]$ and $X_2[n]$, the output of the system will be $Y[n] = Y_1[n] + Y_2[n]$. This is very much like the continuous time system.

We must note that we are adding the entire signals and not just certain points of the signal. Thus, we are basically adding the entire sequence of signals. We refer to the signals as a sequence since in Discrete Systems, we can talk about a point before and a point after the given point.

Proving Additivity

Note that in order to prove that a given system is additive, we should prove that the system is additive for all the points in the sequence.

Thus if we add the points $X_1[n] + X_2[n] + \dots + X_n[n]$, in order to prove the given system is additive, the output has to be $Y[n] = Y_1[n] + Y_2[n] + \dots + Y_n[n]$.

$$\sum X[n] \rightarrow \sum Y[n]$$

However, in order to prove that a given system is not additive, disproving the additivity at even two given points is enough.

The Tax Example

We could now look at the tax example which we had previously seen.

As we know, the tax collected is given by the equation :

$$Y[n] = \alpha.X[n] + \beta.X[n - 1]$$

Here α and β are the rates of taxation at the given instances.

In order to illustrate the additivity of this taxation system, let us consider two states S1 and S2 having the same tax rules. $X_1[n]$ and $X_2[n]$ are the populations of the two states in the nth interval respectively.

Let S be the system that collects the tax. If we consider both the states together, we need to find the tax that is collected together.

Proof

$$X[n] = X_1[n] + X_2[n]$$

$$Y[n] = \alpha.X[n] + \beta.X[n - 1]$$

$$Y[n] = \alpha.X_1[n] + \alpha.X_2[n] + \beta.X_1[n - 1] + \beta.X_2[n - 1]$$

$$Y[n] = [\alpha.X_1[n] + \beta.X_1[n - 1]] + [\alpha.X_2[n] + \beta.X_2[n - 1]]$$

$$Y[n] = Y_1[n] + Y_2[n]$$

The above will be true for any given $X_1[n]$ and $X_2[n]$. We thus conclude that the above system is additive.

1. Introduction to Abstractions

1.9.2. Homogeneity and Shift Invariance.

Introduction

We have looked at the Additive property of Discrete Systems previously. Continuing our discussion on properties, we now look at two new properties of Discrete Systems. Namely, Homogeneity and Shift Invariance.

1.9.3. Definition

The properties of continuous as well as discrete systems seem to be more or less similar. However, there are minor differences in the way we define these properties.

Homogeneity

A Homogeneous system is defined as the one where if the input is scaled by a given constant, the output is also scaled by the same constant.

Shift Invariance

Shift Invariant system is defined as the one where if we shift the input by a given amount of time (or given number of samples in this case), the output will be shifted by the same amount of time (or number of samples).

Note that in discrete systems, we can only shift the input by integral values unlike in continuous systems.

A Homogeneous System

Consider a system with input $X[n]$, giving an output $Y[n]$. Now, we must note that in a discrete memory system, we map the whole input sequence on to the output sequence. It is not a point to point mapping which would make it memoryless.

Thus, when we scale the given input signal by say a constant α , for the system to be homogeneous the output will be scaled by the same constant α . This is true for every possible input $X[n]$ and every possible constant α .

$$X[n] \rightarrow Y[n]$$

$$\alpha.X[n] \rightarrow \alpha.Y[n]$$

Tax example with respect to a Homogeneous System

We can now illustrate the tax system example taken a few sessions before.

If we double the number of people in the state, the tax collected will also be doubled.

Homogeneity insists that the output will be scaled by ‘the very same’ constant with which we scale the input.

A Shift Invariant System

Consider a discrete system, with an input $X[n]$, producing an output $Y[n]$.

Now, if we shift the given input by an integer D , we have the input as $X[n - D]$.

1.9. Lecture 11: Properties of Discrete Systems.

For the system to be shift invariant, it should produce an output which will be shifted by the same integer D.

$$\begin{aligned} X[n] &\rightarrow Y[n] \\ X[n - D] &\rightarrow Y[n - D] \end{aligned}$$

Where $D \in \text{Integer}$ and the above holds true for every D and every $X[n]$.

Tax example with respect to a Shift Invariant System

Going back to the taxation example, what Shift invariance implies is that the rate of the tax remains the same over several years.

So as we jump by a certain amount of time, the tax or the output $Y[n]$ should continue to remain the same.

[inline, color=white]Note that Shift Invariance demands a lot from the system. Shift Invariant Systems are very rare in real life or are probably non-existent. A system could however be Shift Invariant over a long range of period.

1. Introduction to Abstractions

1.9.4. Causality

Introduction

In our discussion on the Discrete Independent Variable systems, we have so far seen 3 properties. Namely Additivity, Homogeneity and Shift Invariance. We now look at another two properties, Causality and Stability. These discussion on the properties of Discrete Systems help us to understand how they are very much similar to the properties of Continuous Systems in nature.

Definition

A system is Causal if the output at anytime depends only on the values of the input at the present time and in the past.

Thus, if two inputs to a Causal System are identical up to some sample n_0 , the corresponding outputs must also be equal up to this point.

A Causal System

A causal System is also referred to as a non-anticipative system since the system output does not anticipate future values of the input.

A real life example of a Causal system could be an RC circuit since a capacitor voltage depends only on the present and past values of the source voltage.

Looking at the definition of a Causal System, we can represent it as follows:

If

$$X_1[n] = X_2[n]$$

upto

$$n \leq n_0$$

, A system is said to be causal if for all $X_1[n], X_2[n], n \leq n_0$

$$Y_1[n] = Y_2[n]$$

1.9.5. Stability

Boundedness

A bounded input can be defined as an input having a bounded maximum magnitude. If we have an input $X[n]$, magnitude of the input will have a maximum value M_x which will always be (strictly) less than ∞ .

The above can be represented as follows:

$$X[n] \leq M_x < \infty$$

Definition

A system is defined to be Stable if every bounded input to the system results in a bounded output over the interval $[n_0, \infty]$.

This must hold for all initial samples. So, as long as we don't input ∞ to our system, we won't get ∞ output.

A Stable System

A Stable System is the one for which small inputs lead to responses that do not diverge. The input of a stable system is always bounded. An example of a Stable System could be as follows:

$$Y[n] = \left(\frac{1}{2T+1}\right) \sum_{k=-T}^T X[n-k]$$

We can see that as $X[n]$ is a bounded value for all values of n by certain number β , then the largest possible magnitude of $Y[n]$ will also be β .

This is because $Y[n]$ is the average of a finite set of values of input.

We looked at a few important properties of Discrete Independent Variable Systems in this Lecture.

1.10. Lecture 12: Impulse Response and Convolution

1.10.1. Introduction

In the previous sessions, we have studied various properties of Discrete Systems. Two of these properties when combined are of great analytical importance and hence of great help. These are, Linearity and Shift Invariance.

In this session, we look at the combination of the above two systems, names as Linear Shift Invariant systems.

1.10.2. A Linear Shift Invariant System

Consider a system "S" which is Linear Shift Invariant. We have already studied that

$$\text{Linearity} = \text{Additivity} + \text{Homogeneity}$$

Thus, a Linear Shift Invariant System will have the properties of:

1. Additivity
2. Homogeneity
3. Shift Invariance

In order to study the various input-output characteristics of a Linear Shift Invariant System, we could apply a certain known input function namely a 'Unit Impulse Response' Sequence.

Unit Impulse Response

An impulse is a sudden, shortlived phenomenon.

A Unit Impulse Response sequence can be defined as the sequence which has just one finite value '1' at the origin and 'zero' else where.

Mathematically it can be written as:

$$\delta[n] = 1, n = 0$$

$$\delta[n] = 0, n \neq 0$$

1. Introduction to Abstractions

Taxation system as a Linear Shift Invariant System

we had earlier seen the taxation system. We had the tax collected as $Y[n]$ given by

$$Y[n] = \alpha.X[n] + \beta.X[n - 1]$$

Additivity

We have already proven that the given system is Additive in the previous session.

Now, in order to find out if the given system is Linear Shift Invariant, we need to check for Homogeneity as well as Shift Invariance.

Homogeneity

In order to prove that the system is Homogenous, we can scale the input $X[n]$ by a constant γ

$$X[n] = \gamma.X[n]$$

Therefore,

$$Y[n] = \alpha.X[n] + \beta.X[n - 1]$$

$$Y[n] = \alpha.\gamma.X[n] + \beta.\gamma.X[n - 1]$$

$$Y[n] = \gamma.(\alpha.X[n] + \beta.X[n - 1])$$

$$Y[n] = \gamma.Y[n]$$

Hence we prove that the Tax system is Homogenous.

Shift Invariance

To prove that the system is Shift Invariant, we Shift the input by an integer 'D'.

$$X[n] = X[n - D]$$

Therefore,

$$Y[n] = \alpha.X[n] + \beta.X[n - 1]$$

$$Y[n] = \alpha.X[n - D] + \beta.X[n - 1 - D]$$

$$Y[n] = Y[n - D]$$

Thus we can see that the Taxation System is

- Additive
- Homogenous
- Shift Invariant

The Taxation System is thus found out to be an example of a 'Linear Shift Invariant' System.

Unit Impulse Response of a Linear Shift Invariant System

Consider a Unit Impulse Sequence given as an input to a Linear Shift Invariant System. Input is given as $\delta[n]$ Therefore,

$$Y[n] = \alpha\delta[n] + \beta\delta[n - 1]$$

Now, we know that $\delta[n] = 1 \text{ at } n = 0$ Hence we find that

$$Y[n] = \alpha$$

$n=0$

$$Y[n] = \beta$$

$n-1=0$ or $n=1$

$$Y[n] = 0$$

$n \neq 0$

Thus the tax collected or the output $Y[n]$ will have only two discrete values.

We thus analyze the unit impulse response of a Linear Shift Invariant System.

1.10.3. Theorem

[inline, color=white!40]The theorem for Discrete Linear Time Invariant System states that 'the unit impulse response characterizes the system completely.'

The above theorem states that given the unit impulse response of a system, it is possible to determine the output $Y[n]$ for an input $X[n]$.

2. The Impulse Response for Discrete LSI Systems

2.0.1. Introduction

The input output relation of a Linear Time Invariant system can be found out if we know its unit impulse response.

We have formally stated this in a theorem that 'the unit impulse response characterizes the system completely.'

We will look at the proof of the above theorem in this session.

2.0.2. Proof of the Theorem

Considering the theorem, we can say that we know that for a given LSI System for a given input $X[n] = \delta[n]$ the output will be $Y[n] = h[n]$.

We shall now look at the proof step by step:

Linear Combination of a Sequence

Consider an input sequence

$$X[n] = [3, 2, -4, 7]$$

Consider 2 as the sample at the origin.

Thus, $X[-1] = 3, X[0] = 2, X[1] = -4, X[2] = 7$.

The above sequence can be written as a combination of unit impulses of different magnitudes.

$$X[n] = 3\delta[n+1] + 2\delta[n] - 4\delta[n-1] + 7\delta[n-2]$$

Therefore,

$$X[n] = X[n+1]\delta[n+1] + X[n]\delta[n] - X[n-1]\delta[n-1] + X[n-2]\delta[n-2]$$

$$X[n] = \sum_{k=-\infty}^{\infty} x[k]\delta[n-k]$$

Shift Invariance

We will now recall the properties of a LSI System. $X[n] = \delta[n]$ Since given system is Shift Invariant, $\delta[n-D] \rightarrow h[n-D]$

Homogeneity

A LSI System is Homogeneous. For a given value of k , the value of $X[k]$ will be constant. scaling the impulse, shifted input by $X[k]$ Thus,

$$X[k]\delta[n-k] \rightarrow X[k]h[n-k]$$

2. The Impulse Response for Discrete LSI Systems

Additivity

LSI System shows the property of Additivity.

$$\sum_{k=-\infty}^{\infty} X[k]\delta[n-k] \rightarrow \sum_{k=-\infty}^{\infty} X[k]h[n-k]$$

The term ' $\sum_{k=-\infty}^{\infty} x[k] * \delta[n - k]$ ' is nothing but $X[n]$ as seen above.

We can now conclude that the output of the system will be definitely known because of the earlier 3 properties discussed.

Therefore,

$$Y[n] = \sum_{k=-\infty}^{\infty} X[k].h[n-k]$$

Hence we prove the theorem "constructively". That is, we now constructively and definitively know the input-output relation once we know the impulse response of the given LSI system.

2.0.3. Introduction to Convolution

The previously obtained output $Y[n]$ is obtained by 'convoluting' $X[n]$ and $h[n - k]$.

This operation that is carried out between the two sequences is known as convolution.

3. Discrete Convolution

3.0.1. Introduction

Earlier we saw the concept of Convolution of two Sequences of Signals.

Convolution is basically an operation between the two sequences which enables us to find out the output sequence.

3.0.2. Definition

Convolution can be strictly defined as the operation between the input signal $X[n]$ and the impulse response $h[n]$ that results in the output $Y[n]$

$$Y[n] = X[n] * h[n]$$

In the above form, the '*' term denotes Convolution. Therefore,

$$Y[n] = \sum_{k=-\infty}^{\infty} X[k].h[n-k]$$

[inline, color=white]Note: The symbol * will henceforth always refer to Convolution of two signals and NOT just simple multiplication.

3.0.3. Procedure for Convolution

Applying the Input Signal

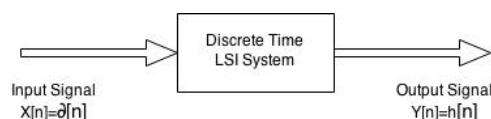


Figure 3.1.: Applied input $X[n] = \delta[n]$

Delaying the Input Signal

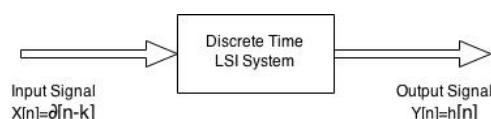


Figure 3.2.: Applied delay to the signal

3. Discrete Convolution

Scaling the delayed Input Signal

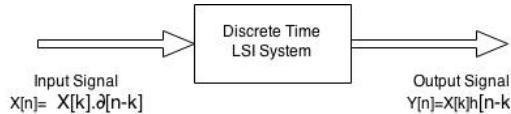


Figure 3.3.: Scaling the input by $X[k]$

Adding the Scaled Impulses over the entire Range

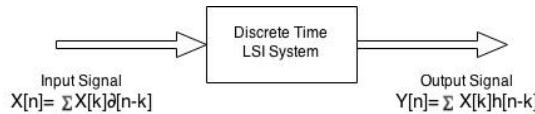


Figure 3.4.: Summing the input over the entire region

3.0.4. Applications

A few Applications of Convolution are listed below:

- To obtain the response (or output) signal of the discrete time signal input applied.
- To obtain overall impulse response signal $h[n]$ for interconnected systems.
- To determine similarity (co-relation) between the transmitted and reflected (or received) signal. This is also called as Auto-Correlation.
- To obtain similarity between 2 different signals. Also called Cross-Correlation.

3.0.5. Examples on Convolution

Example 1:

Calculate the output $Y[1]$ for

$$X[n] = [3, 1, 2] \quad X[0] = 3$$

$$h[n] = [3, 0, 2] \quad X[0] = 3$$

To find out $Y[1]$, replace n by 1 in the Convolution formula.

$$Y[1] = \sum_{k=-\infty}^{\infty} X[k].h[1-k]$$

The signal sequence $h[1 - k]$ will be represented as:

$$h[1 - k] = [2, 0, 3] \quad h[0] = 0$$

While Convoluting 2 sequences, as we have seen in Continuous Time domain, we need to find the part of the signal sequences that overlap.

Therefore,

$$\begin{aligned} Y[1] &= (X[0].h[0]) + (X[1].h[1]) \\ Y[1] &= (3.0) + (1.3) \\ Y[1] &= 3 \end{aligned}$$

Thus we Convolve the signal sequences $X[n]$ and $h[n]$ in order to find the output $Y[n]$ at the point $n = 1$.

3.0.6. Properties of Convolution

Commutativity

If,

$$Y[n] = X[n] * h[n]$$

Then, by the Commutative Property of Convolution,

$$Y[n] = h[n] * X[n]$$

Distributivity

If,

$$Y[n] = X[n] * (h_1[n] + h_2[n])$$

Then, by the Distributive Property of Convolution,

$$Y[n] = X[n] * h_1[n] + X[n] * h_2[n]$$

Associativity

If,

$$Y[n] = X[n] * (h_1[n] * h_2[n])$$

Then, by the Associative Property of Convolution,

$$Y[n] = (X[n] * h_1[n]) * h_2[n]$$

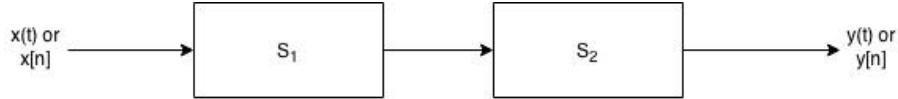
3.1. Lecture 13: Causality of LSI systems

3.1.1. Associativity and Commutativity of LSI Systems

Two more properties of system, namely associativity and commutativity are explored here. These properties are not "new" in the sense that some basic operations like addition and multiplication follow these properties. Note: Both discrete and continuous systems are considered together in the following discussion.

3. Discrete Convolution

Associativity



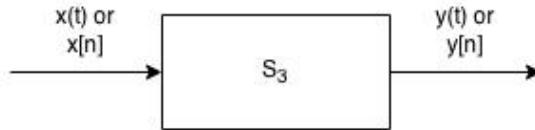
In the above image, we input signal x , either $x(t)$ for continuous system or $x[n]$ for discrete one, to LSI system S_1 (having its impulse response $h_1(t)$ or $h_1[n]$), which can be considered continuous or discrete, depending on the case. The output of system S_1 is fed to LSI system S_2 (having its impulse response $h_2(t)$ or $h_2[n]$), which also can be continuous or discrete. The output of S_2 is y , either $y(t)$ or $y[n]$ in continuous and discrete case respectively. That is, the whole setup can be considered either of the continuous or the discrete case. In the continuous case, both the systems as well as the relevant signals are continuous, whereas in discrete case they all operate on a discrete independent variable. Note here, that the output obtained from the first system is $(x * h_1)$, continuous or discrete. As this signal is fed to the S_2 , we get output as

$$y = (x * h_1) * h_2$$

Now Associativity talks about the equivalence of systems put in series or cascade. Mathematically,

$$(x * h_1) * h_2 = x * (h_1 * h_2)$$

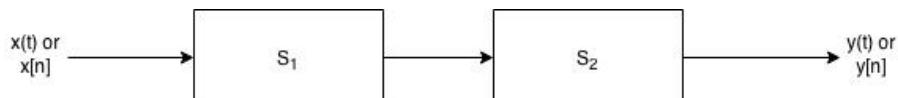
Now if we have a LSI system S_3 , whose impulse response is $h_1 * h_2$, then from above equation it is clear that it will act exactly like these two systems S_1 and S_2 would have acted when applied in series. This is the physical significance of the associativity.



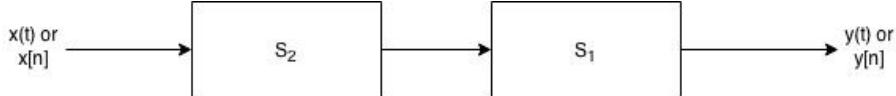
So in short, if we put two systems in a cascade, we can replace them with one single system whose impulse response is equal to the convolution of the impulse responses of these two systems.

Commutativity

Let us apply signal x to system S_1 and then its output to system S_2 , both LSI, to get output as y . Commutativity states that the order in which these systems come is irrelevant to the result i.e. we will get the same output in both the cases. It states that regardless of which of the two systems comes first, the output y is going to be the same.



So the equivalent system which we talked about in the Associativity, is unaffected if the order of these two systems is reversed.



We can also generalize this to n cascaded LSI systems and say that for any permutation of the given n systems, the output is going to be the same. The proof of this is a bit more intricate and is left as an exercise to the student. The proof can be done by using the method of mathematical induction.

What if my systems are not LSI?

All of the above discussion is based on the assumption that all the systems we are considering are LSI in nature. While it may be tempting to generalize these seemingly simple results to all systems, and many times we tend to forget to confirm that the system in question is LSI or not before applying such properties, we must be careful.

It turns out that the commutativity and associativity may not hold if the systems S_1 and S_2 are not both LSI. For example, let's say system S_1 is given by the equation $y(t) = (x(t))^2$ and S_2 by $y(t) = x(t) + 1$. We have taken here continuous variable signals and systems, but the same discussion is valid for the discrete case as well. So if we apply first S_1 and then S_2 , we will get $y(t) = (x(t))^2 + 1$ as the output, while reversing the order, i.e. first applying S_2 and then S_1 would yield $y(t) = (x(t) + 1)^2$. So the both outputs are clearly different.

But this does not mean that we can hurry and conclude that the resultant systems will *never* be the same. Consider, say, S_1 as $y(t) = x(t) + 4$ and S_2 as $y(t) = x(t) - 4$. Now in this case, regardless of which system comes first, we get output $y(t) = x(t)$.

Basically, if even one of the systems is not LSI, we cannot say anything about whether those systems follow associativity or commutativity beforehand: we must examine those systems carefully and determine the answer for the individual case.

3.1.2. Causality of a Linear Shift Invariant System from its Impulse Response

We have noted before that a LSI system can be represented by its impulse response, i.e. we know everything about a system once we know its impulse response. We have seen system properties like causality and stability: according to the previous statement, we should be able to know whether a system is stable and/or causal just by looking at and examining its impulse response. The conditions turn out to be very similar for continuous and discrete variable systems and their respective impulse responses.

Causality of continuous variable LSI systems

We say that a system is causal when its output depends only on the past and the current inputs, and *not* on the future inputs. We know that, for a continuous independent variable system S having impulse response $h(t)$, we can relate its input and output by the equation

$$y(t) = \int_{-\infty}^{+\infty} h(\tau)x(t - \tau) d\tau$$

Note that as τ goes from $-\infty$ to $+\infty$, $t - \tau$ also goes from $-\infty$ to $+\infty$. Note that τ being positive corresponds to the past (as $t - \tau < t$ when τ is positive and ' t ' is the

3. Discrete Convolution

current timeframe) while τ negative to the future and 0 to the current time. Now for the system being causal, we want to eliminate the future inputs, so we must have $h(\tau) = 0$ when τ is negative, so that there's no contribution of the future terms to the $y(t)$, the output. Note that $h(\tau) = 0$ when $\tau < 0$ is sufficient condition for causality of the system. It is easy to prove that this condition is also necessary for the system to be causal i.e. there can be no system which is causal that violates this condition. The proof can be done by using the method of proof by contradiction.

Causality of discrete variable LSI systems

For discrete systems, we have the following equation relating output to the input(s):

$$y[n] = \sum_{\kappa=-\infty}^{+\infty} h[\kappa]x[n - \kappa]$$

Similar to the argument for the continuous case, $\kappa < 0$ corresponds to the future, $\kappa > 0$ to the past and $\kappa = 0$ to the present/current input. So causality will be ensured if all the future inputs are cut off, i.e. $h[\kappa] = 0, \forall \kappa < 0$. This is the sufficient condition for system to be causal. We can also prove the sufficiency by one more way:

Consider signals $x_1[n]$ and $x_2[n]$ such that $x_1[n] = x_2[n] \forall n \leq n_0$.

Note that $x_1[n]$ may not be the same as $x_2[n]$ for any $n \geq n_0$. Also, from given condition (we are assuming the condition we arrived at to reach at the conclusion that the system is causal) $h[\kappa] = 0 \forall \kappa < 0$

Now if the system is causal, we must be able to reach $y_1[n] = y_2[n] \forall n \leq n_0$.

Consider

$$y_1[n] = \sum_{\kappa=0}^{+\infty} h[\kappa]x_1[n - \kappa], n \leq n_0$$

Note that we started the sum from 0 as before zero, in $-\infty$ to 0 region, we already know that $h[\kappa] = 0$. Now observe the summation. it contains $x_1[n - \kappa]$. But as κ is less than 0, we have $n - \kappa > n$, or $x_1[n - \kappa] = x_2[n - \kappa]$.

So the sum can be converted to

$$y_1[n] = \sum_{\kappa=0}^{+\infty} h[\kappa]x_2[n - \kappa]$$

And the right hand side of the equation is essentially $y_2[n]$.

Hence $y_1[n] = y_2[n] \forall n \leq n_0$. Hence the sufficiency of the condition is proved. The necessity can also be proved, by using proof by contradiction.

3.2. Lecture 14: Conditions for Stability

Let us try to see whether system is stable or not, from its impulse response. First we will discuss about the discrete case, and then move on to continuous variable case. In both cases, that a system is stable means it gives a bounded output for every bounded input, or a bounded input signal always yields a bounded signal as an output.

3.2.1. Stability of LSI system from it's impulse response: Discrete variable.

For discrete variable systems, output is related to the input by the following equation:

$$y[n] = \sum_{\kappa=-\infty}^{+\infty} h[\kappa]x[n - \kappa]$$

As we are concerned whether output is bounded or not, we need to look at the modulus of the output, i.e.

$$|y[n]| = \left| \sum_{\kappa=-\infty}^{+\infty} h[\kappa]x[n-\kappa] \right|$$

But we know that, in general,

$$\left| \sum a_i \right| \leq \sum |a_i|$$

So

$$|y[n]| \leq \sum_{\kappa=-\infty}^{+\infty} |h[\kappa]| |x[n-\kappa]|$$

Or,

$$|y[n]| \leq \sum_{\kappa=-\infty}^{+\infty} |h[\kappa]| |x[n-\kappa]|$$

But we know that the input signal is bound, say $x[n] \leq M_x, \forall n$. So

$$|y[n]| \leq M_x \sum_{\kappa=-\infty}^{+\infty} |h[\kappa]|$$

Clearly, if $\sum_{\kappa=-\infty}^{+\infty} |h[\kappa]|$ is finite, then $|y[n]|$ is finite. This is the sufficient condition for a system to be stable. The quantity $\sum_{\kappa=-\infty}^{+\infty} |h[\kappa]|$ is called as the 'absolute sum' of h .

In general, if we have a sequence $\{a_n\}$ then the quantity $\sum_{\kappa=-\infty}^{+\infty} |\{a_n\}|$ is called as the absolute sum of sequence $\{a_n\}$. In short, we take the absolute value of the each term of the sequence, and add those to get the absolute sum of a sequence.

If indeed $h[n]$ is absolutely summable, i.e. its absolute sum is a finite number/absolute sum is convergent, then we can say for some M_h lying between 0 and ∞ ,

$$\sum_{\kappa=-\infty}^{+\infty} |h[\kappa]| \leq M_h$$

So our output in that case becomes

$$|y[n]| \leq M_x M_h$$

This is very important. It means that the sufficiency condition which we have is a constructive condition. It not only tells that the output is bounded, but also talks about its bound.

3.2.2. Stability of LSI system from it's impulse response: Continuous variable.

For continuous variable systems, output is related to the input by the following equation:

$$y(t) = \int_{-\infty}^{+\infty} h(\tau)x(t-\tau) d\tau$$

As we want to discuss about the boundedness of $y(t)$, we will take modulus on both sides:

$$|y(t)| = \left| \int_{-\infty}^{+\infty} h(\tau)x(t-\tau) d\tau \right|$$

3. Discrete Convolution

Using the identity

$$|\int f(x)dx| \leq \int |f(x)|dx$$

We get

$$|y(t)| \leq \int_{-\infty}^{+\infty} |h(\tau)x(t-\tau)|d\tau$$

So

$$|y(t)| \leq \int_{-\infty}^{+\infty} |h(\tau)||x(t-\tau)|d\tau$$

But we know that, as input $x(t)$ is bounded,

$$|x(t)| \leq M_x \quad \forall t, 0 \leq M_x < \infty$$

So

$$|y(t)| \leq \int_{-\infty}^{+\infty} |h(\tau)|M_x d\tau$$

We observe that, $y(t)$ will be bounded if $\int_{-\infty}^{+\infty} |h(\tau)| d\tau$ is bounded. And that is the sufficient condition for a continuous variable system to be stable. Now this integral $\int_{-\infty}^{+\infty} |h(\tau)| d\tau$ is called the 'absolute integral' of $h(\tau)$, similar to the absolute sum.

In this case also, we can estimate the bound on the output as follows:

Let

$$\int_{-\infty}^{+\infty} |h(\tau)| d\tau \leq M_h.$$

Substituting this in the equation we got for the output,

$$|y(t)| \leq M_x M_h$$

Note that, we have discussed here only the sufficiency of the conditions for the stability of systems, both discrete and continuous. We still need to ponder upon whether these conditions are necessary for a system to be stable.

3.2.3. Necessity of the condition for stability

We will now discuss about the necessity of the condition for the stability for discrete variable systems. The proof for continuous variable systems is very similar. We will assume that the output obtained from a system, $y[n]$, when it receives a bounded input is also bounded, i.e.

$$|y[n]| \leq M_y \quad \forall n, 0 \leq M_y < \infty$$

And try to reach at the necessary condition we had earlier got, i.e. try to prove that the absolute sum of $h[n]$ is finite. For that, we will give a specific bounded input to the system, and observe output at a specific point, which should also be bounded if we indeed have a stable system(from our assumption). We know that the input and output for a discrete variable system are related through the following equation:

$$y[n] = \sum_{\kappa=-\infty}^{+\infty} h[\kappa]x[n-\kappa]$$

3.3. Lecture 15: Introduction to Convolution

Now consider that we are looking at the output when n is equal to zero. So the equation now becomes

$$y[0] = \sum_{\kappa=-\infty}^{+\infty} h[\kappa]x[-\kappa]$$

Basically we are trying to bring the absolute sum as an output for some suitable input, which is bounded, and once it is done, we will have proven the necessity of the condition for stability. And the input we choose to give to the system is the following:

$$\begin{aligned} x[n] &= \frac{\overline{h[-n]}}{|h[-n]|}, \text{ if } h[-n] \neq 0 \\ &= 0 \quad , \text{ if } h[-n] = 0 \end{aligned}$$

Where $\overline{h[-n]}$ is the complex conjugate of $h[-n]$. So when we apply it in the equation, it becomes

$$\begin{aligned} n &= -\kappa \quad \text{So} \\ x[-\kappa] &= \frac{\overline{h[\kappa]}}{|h[\kappa]|}, \text{ if } h[\kappa] \neq 0 \\ &= 0 \quad , \text{ if } h[\kappa] = 0 \end{aligned}$$

So we have the equation

$$y[0] = \sum_{\kappa=-\infty}^{+\infty} h[\kappa]x[-\kappa]$$

After substituting $x[-\kappa]$, it becomes

$$\begin{aligned} y[n] &= \sum_{\kappa=-\infty}^{+\infty} \frac{h[\kappa]\overline{h[\kappa]}}{|h[\kappa]|}, \text{ if } h[\kappa] \neq 0 \\ &= 0 \quad , \text{ if } h[\kappa] = 0 \\ &= \sum_{\kappa=-\infty}^{+\infty} |h[\kappa]| \end{aligned}$$

Which is the absolute sum of the impulse response. Now as $y[n]$ is bounded, $y[0]$ must be finite and so must be the absolute sum. Hence we have proved the necessity of the absolute sum of the impulse response being finite for the system being stable. The proof for the continuous variable systems is very similar. It is left to you as an exercise.

3.3. Lecture 15: Introduction to Convolution

3.3.1. Input Output Relationship in LSI Systems: the Impulse Response

We have studied several properties of the systems, both continuous and discrete. One very important concept amongst them is the impulse response for the LSI systems which not only showed us how we can obtain the output from the input for the system, but also the properties of that system. It's a very peculiar relation, which relates the input to the output though. So let us try to study it further: and see how we go about doing that somewhat-complicated looking summation or integral. Here are the expressions, both of

3. Discrete Convolution

summation for the discrete systems and integral for continuous ones, for our reference:
For discrete systems, we'd write:

$$y[n] = \sum_{\kappa=-\infty}^{+\infty} x[\kappa]h[n-\kappa]$$

And for continuous systems,

$$y(t) = \int_{-\infty}^{+\infty} x(\tau)h(t-\tau) d\tau$$

Doesn't it look very similar? The operation we are performing on the input signal x is essentially the same, the only difference is the difference between an integral and a summation, which comes from the nature of the independent variable: whether that variable is discrete or continuous. So let's examine this operation carefully. How do we actually perform it? Let's take the simpler version of the summation first:

Consider following $x[n]$:

$$\begin{aligned} x[n] &= -1, & n = -1 \\ &= 9, & n = 0 \\ &= 5, & n = 1 \\ &= 3, & n = 2 \\ &= 0, & \text{otherwise} \end{aligned}$$

and $x[n]$ is zero elsewhere.

Arrow Notation for writing $x[n]$ Conveniently

We have represented $x[n]$ above, but this way seems a little bit tedious. So we introduce a new neat notation to write finite-length discrete signals, called the arrow notation. We basically write out all of the values $x[n]$ takes, as n changes, and indicate n corresponding to any one of the values. So we'd write the same signal in arrow notation as:

$$\begin{matrix} x[n] = & -1 & 9 & 5 & 3 \\ & \uparrow & & & \\ & & 0 & & \end{matrix}$$

Now instead of zero, it may as well have been any other number. For example,

$$\begin{matrix} x[n] = & -1 & 9 & 5 & 3 \\ & \uparrow & & & \\ & & -1 & & \end{matrix}$$

is an equally valid expression. The output is assumed to be zero everywhere else, at the points which are not shown.

3.3.2. Getting the output for a discrete system

So, back to our finding the output: we have the $x[n]$:

$$\begin{matrix} x[n] = & -1 & 9 & 5 & 3 \\ & \uparrow & & & \\ & & 0 & & \end{matrix}$$

And $h[n]$:

3.3. Lecture 15: Introduction to Convolution

$$h[n] = \begin{matrix} 1 & -1 & 2 \\ \uparrow \\ 0 \end{matrix}$$

We want to find $y[n]$, where

$$y[n] = \sum_{\kappa=-\infty}^{+\infty} x[\kappa]h[n-\kappa]$$

Let us, for simplicity, take a specific output, say at $n = 0$. Hence our equation becomes

$$y[0] = \sum_{\kappa=-\infty}^{+\infty} x[\kappa]h[-\kappa]$$

So for every κ , we need the value of $x[\kappa]$ and $h[-\kappa]$. Let us tabulate κ , $x[\kappa]$ and $h[-\kappa]$ for relevant κ 's:

κ	2	-1	0	1	2
$x[\kappa]$	0	-1	9	5	3
$h[-\kappa]$	2	-1	1	0	0

So effectively,

$$\begin{aligned} y[0] &= \sum_{\kappa=-\infty}^{+\infty} x[\kappa]h[-\kappa] \\ &= \sum_{\kappa=-1}^0 x[\kappa]h[-\kappa] \\ &= (-1 \times -1) + (9 \times 1) = 10 \end{aligned}$$

Hence we get $y[0] = 10$. Similarly we can $y[n]$ for other values of n .

So we have seen how to calculate the output $y[n]$ for 0. Now let's just do that for some other point, in case we miss some thing because of the simplicity zero sometimes offers, and then we'll try to generalize it.

If we observe, ' n ' affects the $h[n-\kappa]$ term, and $x[\kappa]$ is independent of it, for some particular κ . Now $h[n-\kappa]$ becomes zero when $n = \kappa$. If κ is after n , which means $\kappa > n$, we have the indices before zero going into the $h[n-\kappa]$ term. If κ is before n , the situation is exact reverse, and we have indices after zero going in that term. Let's say if $n = 2$, we have our 0 point at $\kappa = n = 2$. The values of κ after 2 correspond to the input to the impulse response term being -1, -2 and so on, and the values of κ before 2 correspond to the input to the impulse response term being 1, 2 and so on. So now we have $h[2-\kappa]$ as

$$h[2-\kappa] = \begin{matrix} 1 & -1 & 2 \\ \uparrow \\ \kappa & & 2 \end{matrix}$$

Now compare this with $h[0-\kappa]$ we had in the last calculation:

$$h[-\kappa] = \begin{matrix} 1 & -1 & 2 \\ \uparrow \\ \kappa & & 0 \end{matrix}$$

Now we can easily calculate $h[2-n]$, like we did the last time, and it turns out to be 16 (Confirm by doing it yourself!). So that is our $y[2]$.

So for a given n , we first put down whatever is at 0, i.e. $\kappa = n$ point. Then we put those terms which are after zero sequentially before n i.e. 1 at $n-1$, 2 at $n-2$ and so on; and

3. Discrete Convolution

we put those terms which are before zero sequentially after n i.e. -1 at $n + 1$, -2 at $n + 2$ and so on.

We can also understand mathematically how this comes about. Considering n fixed and κ as the our variable, consider that we have a certain $h[\kappa]$. Now $h[\kappa + n]$ is nothing but $h[\kappa]$ shifted forward by n . Now when we replace κ by $-\kappa$, we will have our signal reflect about zero i.e. its value at a go to $-a$ for all integers a . So the zero point, for example, which was at $\kappa = -n$ in $h[\kappa + n]$, now lies at $\kappa = n$ in $h[-\kappa + n]$. All other points are put in the same order with respect to this zero, but to the opposite side than they were before.

3.3.3. Getting the Output for the General Case: the Train Platform Analogy

We have seen how $h[n - \kappa]$ is constructed. First, $h[0]$ is brought to $\kappa = n$. Then we bring $h[1], h[2]$ etc. to $\kappa = n - 1, n - 2$ etc. respectively; and $h[-1], h[-2]$ etc. to $\kappa = n + 1, n + 2$ etc. respectively. We do this all for a fixed n . If we change the value of n , say by +1, the whole structure will move forward by 1 step. However, the internal structure, the order in which particular outputs come is still the same, it remains retained. So we don't have to worry about constructing $h[n - \kappa]$ again and again: we can just create it once for, say $n = 0$, and then as the n changes, shift it accordingly. This eases our calculations, rather the representation very much: and computing the output seems less daunting than it did, for we have a proper method now telling us how to go about doing that.

We can think of it as being similar to a train and a platform. $x[\kappa]$ is the platform, which remains fixed on the κ -scale, and $h[n - \kappa]$ is the train, moving according to n . Basically we are seeing where the zero is falling for some n , and then moving the whole 'train' of $h[n - \kappa]$ accordingly, so that point can be considered as the engine which drives the train. And the other values like $h[1], h[2]$ are like bogies attached to that engine(although it seems rather weird that they are attached on the both sides of an engine, but leaving that aside). So if we have a nonzero value of, say $h[4]$, then we can say that there's a person there. Similarly, at position κ , if $x[\kappa] \neq 0$, we can say that there's a person standing there on the platform. And When there's a person both inside the bogie and on the platform at the same position(of κ), they both shake their hands and the numbers(values of both $x[n - \kappa]$ & $h[\kappa]$) get multiplied. The combined effect of all such multiplications is the value of output for a particular n .

Let's look at the formula for computing the output $y[n]$:

$$y[n] = \sum_{\kappa=-\infty}^{+\infty} x[\kappa]h[n - \kappa]$$

Here, $x[\kappa]$ represents the person on the platform at the location κ , $h[n - \kappa]$ represents the person in the bogie or coach who is currently at that position. So they both shake hands and the values of both of these get multiplied. The sigma of the summation represents that we are taking into consideration the combined effect of all these handshakes.

Well, using the tools we just developed, let us write the complete output for the example we have considered here:

$$\begin{aligned} x[n] &= -1 & 9 & 5 & 3 \\ &\uparrow \\ &0 \\ h[n] &= 1 & -1 & 2 \\ &\uparrow \\ &0 \end{aligned}$$

So we have

$$\begin{array}{c} x[\kappa] = -1 \quad 9 \quad 5 \quad 3 \\ \uparrow \\ \kappa \quad 0 \\ h[n-\kappa] = 1 \quad -1 \quad 2 \\ \uparrow \\ \kappa \quad n \end{array}$$

Thus, by putting different values of n, we get

n	y[n]
$n \leq -2$	= 0
$n = -1$	$= (-1) \times 1 = -1$
$n = 0$	$= 1 \times 9 + (-1) \times (-1) = 10$
$n = 1$	$= 1 \times 5 + (-1) \times 9 + 2 \times (-1) = -6$
$n = 2$	$= 1 \times 3 + (-1) \times 5 + 2 \times 9 = 16$
$n = 3$	$= (-1) \times 3 + 2 \times 5 = 7$
$n = 4$	$= 2 \times 3 = 6$
$n \geq 5$	= 0

Now this process by which we computed the output given the input and the impulse response has a name: it is called "**convolution**" and we say that we get the output by convolving $x[n]$ and $h[n]$. So in general, when we say $a[n]$ convolves with $b[n]$ to give $c[n]$ we mean

$$c[n] = \sum_{\kappa=-\infty}^{+\infty} a[\kappa] b[n-\kappa]$$

It is denoted as

$$c[n] = a[n] * b[n] \quad (3.1)$$

or

$$c[n] = (a * b)[n] \quad (3.2)$$

We will learn more about this convolution in subsequent lectures.

3.4. Lecture 16: The Convolution Operator

3.4.1. Introduction

In the previous lecture, we used the Train-Platform analogy, to look at discrete convolutions. In this lecture, we will extend this analogy to continuous independent variable convolutions. We will also take some examples to illustrate the concept. We'll then look at some properties of the convolution operator in the next lecture.

3.4.2. Convolution of Continuous Signals

In the previous lecture, we defined the convolution of two discrete signals $x[n]$ and $y[n]$ as:

$$(x * y)[n] = \sum_{k=-\infty}^{+\infty} x[k] y[n-k] \quad (3.3)$$

If we want to write a similar expression for continuous signals, we just replace the summation by an integration. The convolution of two continuous signals $x(t)$ and $y(t)$ is:

$$(x * y)(t) = \int_{-\infty}^{+\infty} x(\tau) y(t-\tau) d\tau \quad (3.4)$$

3. Discrete Convolution

3.4.3. Convolution of continuous signals – Example

Let's take an example to illustrate this concept. Consider a pulse: (the platform)

$$x_2(t) = B [u(t - t_2) - u(t - (t_2 + T_2))]$$

being convolved with: (the train)

$$x_1(t) = A [u(t - t_1) - u(t - (t_1 + T_1))]$$

(where $u(t)$ is the unit step function)

Note that x_1 and x_2 are just pulses of height A and B , starting at $t = t_1$ and $t = t_2$, and of length T_1 and T_2 , respectively. Assume, for the moment, that $T_2 > T_1$ (We will soon show that this assumption can be made without loss of generality, when we prove Commutativity of Convolution). Now, we know from Equation 3.4 that

$$(x_2 * x_1)(t) = \int_{\tau=-\infty}^{+\infty} x_2(\tau) x_1(t - \tau) d\tau$$

Think of τ as the indexing on the platform. Thus, we can say that the platform starts at $\tau = t_2$ and goes up till $\tau = t_2 + T_2$, with people uniformly distributed over the whole platform.

Looking at the train, $x_1(t - \tau) = x_1(t_1)$ when $\tau = t - t_1$ and $x_1(t - \tau) = x_1(t_1 + T_1)$ when $\tau = t - (t_1 + T_1)$. Thus, the ends of the train are at $\tau = t - t_1$ and $\tau = t - (t_1 + T_1)$. Until the "front end of the train" reaches the "platform" – this happens when $\tau = t - t_1 = t_2$ – the value of $(x_2 * x_1)(t)$ (the number of handshakes) is 0. Similarly, after the "rear end of the train" has left the "platform" – this happens when $\tau = t - (t_1 + T_1) = t_2 + T_2$ – the value is 0 again.

In between these two extremes, the value of $(x_2 * x_1)(t)$ essentially depends only on the amount of "overlap" between the platform and the train. In other words, it depends on how much of the train is on the platform. It can be shown that:

$$(x_2 * x_1)(t) = \begin{cases} 0 & , \text{ if } t \leq t_1 + t_2 \\ AB(t - t_1 - t_2) & , \text{ if } t_1 + t_2 < t \leq (t_1 + T_1) + t_2 \\ ABT_1 & , \text{ if } (t_1 + T_1) + t_2 < t \leq t_1 + (t_2 + T_2) \\ AB((t_2 + T_2) + (t_1 + T_1) - t) & , \text{ if } t_1 + (t_2 + T_2) < t \leq (t_1 + T_1) + (t_2 + T_2) \\ 0 & , \text{ if } (t_1 + T_1) + (t_2 + T_2) < t \end{cases}$$

(train hasn't reached platform)
(Region 1: train is entering the platform)
(Region 2: train is fully on the platform)
(Region 3: train is leaving the platform)
(train has left the platform)

3.5. Lecture 17: Properties of the Convolution Operator

3.5.1. Introduction

In the previous lecture, we looked at convolution, we also studied an example of the convolution of two very simple signals in detail. Here, we will look at two of the major properties of this operator, namely Commutativity and Associativity.

3.5.2. Commutativity

A binary operator (any operator that takes two operands and produces a result) is said to be commutative if changing the order of the operands does not change the result. For example, addition is commutative (e.g. $5 + 3 = 8 = 3 + 5$), but subtraction is not (e.g. $5 - 3 \neq 3 - 5$).

We are now in a position to prove the commutativity of convolution:

Theorem 1. *Convolution is commutative, $x * y = y * x$.*

Proof.

Case 1.1. *Signals are discrete*

The convolution of two discrete signals $x[n]$ and $y[n]$ is defined as:

$$(x * y)[n] = \sum_{k=-\infty}^{+\infty} x[k] y[n-k] \quad (3.5)$$

Note that this equation is true $\forall n$. Substitute $n - k = l$ (which also implies $k = n - l$) in Equation 3.5. For any given n , as k takes all the integral values from $-\infty$ to $+\infty$, $l = n - k$ also takes all the integral values, albeit in reverse order. Since the order in which we add the terms does not matter, we get:

$$(x * y)[n] = \sum_{l=-\infty}^{+\infty} x[n-l] y[l] \quad (3.6)$$

which, if you look carefully at it, is really $(y * x)[n]!$

Case 1.2. *Signals are continuous*

The convolution of two continuous signals $x(t)$ and $y(t)$ is:

$$(x * y)(t) = \int_{\tau=-\infty}^{+\infty} x(\tau) y(t-\tau) d\tau \quad (3.7)$$

Substitute $t - \tau = \alpha$. Here, the order does matter. For any fixed t , as τ varies from $-\infty$ to $+\infty$, $\alpha = t - \tau$ varies from $+\infty$ to $-\infty$. Also, since $\alpha = t - \tau$, $d\alpha = -d\tau$. This gives us:

$$(x * y)(t) = \int_{\alpha=+\infty}^{-\infty} x(t-\alpha) y(\alpha) (-d\alpha) \quad (3.8)$$

By reversing the order of the order of integration, we get rid of the $-$ sign. Thus:

$$(x * y)(t) = \int_{\alpha=-\infty}^{+\infty} x(t-\alpha) y(\alpha) d\alpha \quad (3.9)$$

This is nothing but $(y * x)(t)$. □

3. Discrete Convolution

3.5.3. Associativity

A binary operator is said to be associative if changing the order in which the operation is done on a sequence of (2 or more) operands (keeping the order of operands fixed), does not change the result. For example, multiplication is associative: $(2 \times 3) \times 5 = 30 = 2 \times (3 \times 5)$, but division is not: $\left[\frac{\frac{2}{4}}{5} = \frac{1}{10} \right] \neq \left[\frac{2}{\frac{4}{5}} = \frac{5}{2} \right]$.

We can now prove that convolution is associative

Theorem 2. *Convolution is Associative, $(x * y) * z = x * (y * z)$*

Proof.

Case 2.1. *Signals are discrete*

$$\begin{aligned} ((x * y) * z)[n] &= \sum_{l=-\infty}^{+\infty} (x * y)[l] z[n - l] \\ &= \sum_{l=-\infty}^{+\infty} z[n - l] \left(\sum_{k=-\infty}^{+\infty} x[k] y[l - k] \right) \\ &= \sum_{k,l=-\infty}^{+\infty} x[k] y[l - k] z[n - l] \end{aligned} \tag{3.10}$$

(This is true because l and k vary independently)

(Substitute $k = l_1$, $l - k = k_1 \implies n - l = n - l_1 - k_1$)

$$\begin{aligned} &= \sum_{l_1,k_1=-\infty}^{+\infty} x[l_1] y[k_1] z[n - l_1 - k_1] \\ &= \sum_{l_1=-\infty}^{+\infty} x[l_1] \left(\sum_{k_1=-\infty}^{+\infty} y[k_1] z[(n - l_1) - k_1] \right) \end{aligned}$$

(This is true because l_1 and k_1 vary independently)

$$\begin{aligned} &= \sum_{l_1=-\infty}^{+\infty} x[l_1] (y * z)[n - l_1] \\ &= (x * (y * z))[n] \end{aligned} \tag{3.11}$$

Case 2.2. Signals are continuous

$$\begin{aligned}
 ((x * y) * z)(t) &= \int_{\alpha=-\infty}^{+\infty} (x * y)(\alpha) z(t - \alpha) d\alpha \\
 &= \int_{\alpha=-\infty}^{+\infty} z(t - \alpha) \left(\int_{\lambda=-\infty}^{+\infty} x(\lambda) y(\alpha - \lambda) d\lambda \right) d\alpha \\
 &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x(\lambda) y(\alpha - \lambda) z(t - \alpha) d\lambda d\alpha
 \end{aligned} \tag{3.12}$$

This is an integration over 2 dimensions. Perform coordinate transformation:

$\begin{bmatrix} \alpha_1 \\ \lambda_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \lambda \\ \alpha \end{bmatrix}$, ($\alpha_1 = \lambda$, $\lambda_1 = \alpha - \lambda$). The limits don't change, here.

The Jacobian of this transformation is $\begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = 1 \implies d\lambda d\alpha = 1 \times d\alpha_1 d\lambda_1$

$$\begin{aligned}
 &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x(\alpha_1) y(\lambda_1) z(t - \lambda_1 - \alpha_1) d\alpha_1 d\lambda_1 \\
 &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x(\alpha_1) y(\lambda_1) z(t - \lambda_1 - \alpha_1) d\lambda_1 d\alpha_1 \quad (\text{By Fubini's Theorem}) \\
 &= \int_{\alpha_1=-\infty}^{+\infty} x(\alpha_1) \left(\int_{\lambda_1=-\infty}^{+\infty} y(\lambda_1) z((t - \alpha_1) - \lambda_1) d\lambda_1 \right) d\alpha_1
 \end{aligned}$$

(This is true because α_1 and λ_1 vary independently)

$$\begin{aligned}
 &= \int_{\alpha_1=-\infty}^{+\infty} x(\alpha_1) (y * z)(t - \alpha_1) d\alpha_1 \\
 &= (x * (y * z))(t)
 \end{aligned} \tag{3.13}$$

□

3.5.4. Consequences of Commutativity and Associativity of Convolution

Consider n LSI systems $S_1, S_2, S_3, \dots, S_n$, having impulse responses $h_1, h_2, h_3, \dots, h_n$, respectively. Suppose we cascade these n systems together, i.e. we give the input x to S_1 , take its output and give it to S_2 , then take its output and give it to S_3 and so on and so forth until we finally give the output of S_{n-1} to S_n . We call the output of S_n , the "final" output, y . By definition,

$$y = (((((x * h_1) * h_2) * h_3) \cdots * h_n)) \tag{3.14}$$

However, using associativity repeatedly, we can show that

$$y = (x * (((h_1 * h_2) * h_3) \cdots * h_n))$$

Once this has been done, we use commutativity and associativity repeatedly and show that,

$$(((h_1 * h_2) * h_3) \cdots * h_n) = (((h_{i_1} * h_{i_2}) * h_{i_3}) \cdots * h_{i_n})$$

3. Discrete Convolution

where $h_{i_1}, h_{i_2}, h_{i_3}, \dots, h_{i_n}$ is any permutation of $h_1, h_2, h_3, \dots, h_n$. This proves the fact that the cascade of these n systems can be replaced by a single system having an impulse response

$$h \equiv (h_1 * h_2 * h_3 \cdots * h_n) \quad (3.15)$$

Here, the expression on the right hand side makes sense only because we have proved that a series of convolutions can be computed in any order. This is a very profound statement, because it means that the order in which the n systems are cascaded does not affect the output. The output will depend only on the input, and not on the order in which the systems are arranged.

3.6. Lecture 18: Conclusion of Module 1 & Summary

3.6.1. Introduction

We have studied quite a bit about Linear Shift Invariant (LSI) Systems in the previous lectures. In this lecture, we will look at what the typical Input-Output relation of an LSI system looks like.

3.6.2. Typical System Description of LSI System

If you recall, the system description of the RC circuit looked like

$$y(t) + RC \frac{d}{dt} y(t) = x(t)$$

We would like to know what the input-output relation of any LSI system looks like, in general.

It turns out that many (continuous variable) LSI systems can be characterized by an equation of the form

$$\begin{aligned} y(t) = & A_1 \frac{dy(t)}{dt} + A_2 \frac{d^2y(t)}{dt^2} + \cdots + A_N \frac{d^N y(t)}{dt^N} \\ & + B_0 x(t) + B_1 \frac{dx(t)}{dt} + B_2 \frac{d^2x(t)}{dt^2} + \cdots + B_M \frac{d^M x(t)}{dt^M} \end{aligned} \quad (3.16)$$

where the coefficients $A_1, A_2, \dots, A_N, B_0, B_1, B_2, \dots, B_M$ are all constants. Also, this equation is linear in all terms. Therefore, this form is called a **Linear Constant Coefficient Differential Equation (LCCDE)**

Similarly, many discrete LSI systems can be characterized by¹:

$$\begin{aligned} y[n] = & A_1 y[n-1] + A_2 y[n-2] + \cdots + A_N y[n-N] \\ & + B_0 x[n] + B_1 x[n-1] + B_2 x[n-2] + \cdots + B_M x[n-M] \end{aligned} \quad (3.17)$$

This form is called a **Linear Constant Coefficient Difference Equation**. This too is abbreviated as **LCCDE**.

We will study these LCCDEs in greater detail when we study Fourier Transforms in subsequent lectures.

¹Here we have implicitly assumed that the system is causal, but there's no reason to do so. You can have terms like $B_{-1}x[n+1]$.

3.6.3. Summary of Module One

In this module, we have looked at Discrete and Continuous Variable systems as abstractions of real-world systems. We have also studied some properties of these systems:

1. Additivity
2. Homogeneity
3. Shift Invariance
4. Memory
5. Causality
6. Stability (BIBO Stability)

We then saw how Linearity (Additivity AND Homogeneity) and Shift Invariance made systems “nice” to study. We then looked at LSI systems in detail – we observed that for LSI systems, we need only study the output of the system when given a special input, namely the unit impulse. We studied the unit impulse in both Continuous and Discrete contexts. We then studied the *response* to this impulse, which is called the unit impulse response, in some detail. We also saw how we could predict the output of the system to any given input, using convolution.

We then tried to understand the properties of these systems in terms of this impulse response. We saw that an LSI system is causal iff² the impulse response is absent until $t = 0$. We also saw how absolute summability (for discrete systems) or absolute integrability (for continuous systems) of the impulse response guarantees the BIBO Stability of the LSI system.

With this, we conclude module 1 of EE210.1X. In the next month or so, we will study signals and systems in a different “transformed” domain, called the frequency domain, using the Fourier Transform.

²This is not a typo! **iff** stands for "if and only if".

Part II.

Frequency Domain

4. Sinusoids and LSI Systems

4.1. Module 2: Lecture 1 Sinusoidal Signals and their Properties

Introduction

In the first module, we have been looking at signals and systems in what is called their natural domain. By natural domain, we mean that the independent variable of the signal is the same as it was when the signal was recorded or observed or instituted. An example of a signal in its natural domain is a speech signal expressed as a function of time, as time is the naturally associated independent variable with the speech signal.

For processing signals, we need appropriate systems. Now, what processing needs to be done by the system can be a tricky thing to specify in the natural domain.

For example, say we want to separate the voices of male and female singers from an audio recording of a chorus. Although it is qualitatively easier to understand, one can't give the description of such a system in the time domain. We need to have a broader view of systems, and more importantly, of signals, to be able to tackle this problem.

So in this module we are going to take the first step towards a change of paradigm, a change of world-view in describing signals and systems. In this lecture we will learn how to deal with phase changes in sinusoidal signals and pass sinusoidal signals into stable LSI systems in an intelligent and insightful manner. We will look into why we introduced complex signals in our previous analysis and how these tools make the mathematical modelling simple. We then go on to understanding how to deal with signals beyond the domain of sinusoids. We begin by understanding which signals are special from the point of view of the signals and systems. Let's visit the ideal voltage generator for that purpose.

4.1.1. The Ideal Voltage generator

The ideal voltage generator consists of a circular conducting coil rotating in a constant magnetic field.

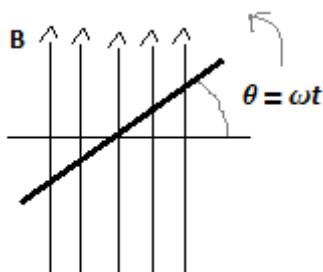


Figure 4.1.: Voltage generator (side view)

Let the area enclosed by the coil be A , the angular frequency of rotation be ω such that

4. Sinusoids and LSI Systems

$\theta = \omega t$, and the value of the constant magnetic field be B . Now, by Faraday's law, the electric field \mathcal{E} is defined as

$$\mathcal{E} = -\frac{\partial \Phi_B}{\partial t}$$

where the flux Φ_B in this case is given by

$$\Phi_B = BA \cos \theta = BA \cos \omega t$$

Hence the electric field will be

$$\mathcal{E} = -BA \omega \sin \omega t$$

This is the principle of generation of the voltage supply that we receive at our residences. This is one important reason why the sinusoidal voltage is highly favoured and deeply studied. And of course there are many other reasons why. Let us have a look at some interesting properties of sinusoids.

4.1.2. Properties of sinusoids

The first interesting thing about sine waves is that when you add two sinusoids of the same frequency, it gives you back another sinusoid of the *same* frequency. Let's prove this formally. Say we have two sinusoids, $x_1(t)$ and $x_2(t)$, of the same angular frequency Ω , given by

$$\begin{aligned}x_1(t) &= A_1 \cos(\Omega t + \phi_1) \\x_2(t) &= A_2 \cos(\Omega t + \phi_2)\end{aligned}$$

Now consider a linear combination of the two sinusoids,

$$x(t) = \alpha_1 x_1(t) + \alpha_2 x_2(t)$$

Note that we can write x_1 and x_2 as

$$\begin{aligned}x_1(t) &= A_1 \{ \cos \Omega t \cos \phi_1 - \sin \Omega t \sin \phi_1 \} \\x_2(t) &= A_2 \{ \cos \Omega t \cos \phi_2 - \sin \Omega t \sin \phi_2 \}\end{aligned}$$

Hence we can write $x(t)$ as

$$x(t) = [\alpha_1 A_1 \cos \phi_1 + \alpha_2 A_2 \cos \phi_2] \cos \Omega t + [-\alpha_1 A_1 \sin \phi_1 - \alpha_2 A_2 \sin \phi_2] \sin \Omega t$$

Writing the two time-independent coefficients as P and Q , we get

$$x(t) = P \cos \Omega t + Q \sin \Omega t = \sqrt{P^2 + Q^2} \left\{ \frac{P}{\sqrt{P^2 + Q^2}} \cos \Omega t + \frac{Q}{\sqrt{P^2 + Q^2}} \sin \Omega t \right\}$$

This can be easily seen to be

$$x(t) = \sqrt{P^2 + Q^2} \cos(\Omega t + \Phi)$$

with

$$\Phi = -\tan^{-1} \left(\frac{Q}{P} \right)$$

Hence, linearly combining two sinusoids of the same frequency gives back a sinusoid with the same frequency.

Another interesting property is that when we differentiate a sine wave, we get back a sinusoid of the same frequency.

$$\frac{d}{dt}\{A \cos(\Omega t + \phi)\} = -A\Omega \sin(\Omega t + \phi)$$

It doesn't matter whether it is cosine or sine since they differ only by a phase of $\pi/2$.

$$\cos(\Omega t + \phi) = \sin(\Omega t + \phi + \pi/2)$$

Now let's see how we can describe a change in amplitude of a sinusoid. Suppose the original sinusoid is given by

$$x_1(t) = A \cos(\Omega t + \phi)$$

and let the changed sinusoid is given by

$$x_2(t) = B \cos(\Omega t + \phi)$$

where A and B are positive constants. Then,

$$x_2(t) = \frac{B}{A} x_1(t)$$

So the change of amplitude in a sinusoid is simply described by a multiplying constant. But this is not true for a phase change. Suppose the original sinusoid is given by

$$x_1(t) = A \cos(\Omega t + \phi_1)$$

and the changed sinusoid is given by

$$x_2(t) = A \cos(\Omega t + \phi_2)$$

Unlike the amplitude case, $x_2(t)/x_1(t)$ doesn't turn out to be a constant independent of time. This will cause a problem while dealing with inductors or capacitors. In inductors, for example, the voltage drop is proportional to the derivative of the current passing through it. Hence, if a sinusoidal current is passing through it, the voltage drop across it will also be a sinusoid, but will have a phase difference of 90° with the current. Hence, we cannot establish a relationship like the resistor ($V = RI$) in case of the inductor since V/I won't be a constant independent of time.

To deal with this problem, we will have to use the rotating complex number representation of the sinusoid.

4.1.3. Sinusoids as Rotating Complex Numbers

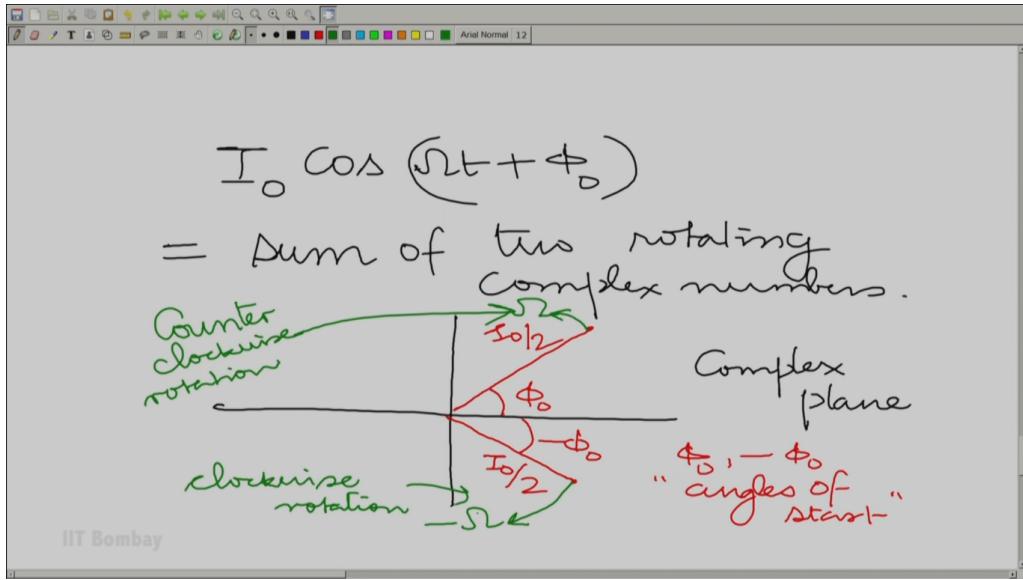
We can think of a sinusoid $I_0 \cos(\Omega t + \phi_0)$ as a combination of two rotating complex numbers. Let the first begin at time $t = 0$ with an angle of ϕ_0 . Let it have a magnitude of $I_0/2$ and let the other begin with the same magnitude $I_0/2$ but the opposite starting angle $(-\phi_0)$. Let them both rotate with an angular velocity of Ω , but the first one in counter clockwise and the second one in clockwise direction.

We can express these two complex numbers in the polar form as

$$c_1 = \frac{I_0}{2} e^{j\{\Omega t + \phi_0\}}$$

$$c_2 = \frac{I_0}{2} e^{-j\{\Omega t + \phi_0\}}$$

4. Sinusoids and LSI Systems



where $j = \sqrt{-1}$. One can see that $c_1 + c_2$ gives back the original sinusoid.

It seems like a silly thing to do to describe a sinusoid as a combination of two rotating complex numbers. But the reason for doing this is clear when we look at the inductor once again.

Let's assume for now that we provide to the inductor, an input which is just one of those rotating complex numbers. So

$$I = \frac{I_0}{2} e^{j\{\Omega t + \phi_0\}}$$

Now, the voltage across an inductor is given by

$$V = L \frac{dI}{dt} = L \frac{I_0}{2} (j\Omega) e^{j\{\Omega t + \phi_0\}}$$

Hence we can see that

$$V/I = j\Omega L$$

Now the ratio of the voltage and current for an inductor is a complex *constant independent of time!*

Similarly, for the other rotating complex number, we will get V/I again to be a constant, but with a minus sign ($-j\Omega L$). So in some sense the actual angular frequency, whether positive or negative, is reflected. It is left as an exercise to do a similar analysis for capacitances and verify that for capacitors,

$$V/I = \frac{1}{j\Omega C}$$

This makes the analysis of RLC circuits (circuits consisting of Resistor(s), Inductor(s) and Capacitors(s)) much easier as we can interpret these constants as generalised resistances, or *impedances*. In that sense, the resistor has an impedance of R , an inductor has an inductance of $j\Omega L$ and a capacitor has an impedance of $1/j\Omega C$.

Notice that c_1 and c_2 defined above are complex conjugates of each other. So whatever happens to c_1 is mirrored in c_2 . For example, the V/I for c_2 ($-j\Omega L$) is the complex conjugate of the V/I for c_1 ($j\Omega L$).

The analysis of sinusoids using rotating complex numbers is known as ‘phasor’ analysis.

As the rotating complex numbers are constant in magnitude, but change their phase, they are called ‘phasors’.

This is the reason why we earlier demanded that our systems should accept a complex signal in general.

In a nutshell, dealing with amplitudes is easy but the dealing with phases is a problem in sinusoids and that problem is overcome when you go to the phasor instead of the corresponding sinusoid. Let us see how a general LSI system responds to a phasor or sinusoid as an input.

4.2. Module 2: Lecture 2

Sinusoidal Input to LSI Systems

4.2.1. Understanding Phase Change In Sinusoids and Phasors

We write a sinusoid as a combination of two rotating complex numbers, which we will henceforth call phasors, moving in opposite directions.

$$2A\cos(\Omega t + \phi) = Ae^{j(\Omega t + \phi)} + Ae^{-j(\Omega t + \phi)}$$

Let us focus our attention on one of the phasors for now. Let us look at what happens when we introduce a phase change.

$$\begin{aligned} Ae^{j(\Omega t + \phi)} &\xrightarrow{\text{Phase Change}} Ae^{j(\Omega t + \phi + \Delta\phi)} \\ Ae^{j(\Omega t + \phi + \Delta\phi)} &= Ae^{j(\Omega t + \phi)} e^{j\Delta\phi} \end{aligned}$$

We hence find that a change of phase results just in a multiplying factor which is a constant independent of time.

Now we try to do a similar calculation for the complex conjugate (the phasor moving in the opposite direction).

$$Ae^{-j(\Omega t + \phi)} \xrightarrow{\text{Phase Change}} Ae^{-j(\Omega t + \phi)} e^{-j\Delta\phi}$$

Notice that the multiplying constant in this case is the complex conjugate of the previous multiplying factor we found.

$$2A\cos(\Omega t + \phi) \xrightarrow{\text{Phase Change}} Ae^{j(\Omega t + \phi)} e^{j\Delta\phi} + Ae^{-j(\Omega t + \phi)} e^{-j\Delta\phi}$$

We hence understand that using only one phasor is enough to fully determine the behaviour of the sinusoid. This explains why we used complex signals in our previous analysis, due to their ease in mathematical manipulation.

We will now see the result of inputting a sinusoidal signal to a stable linear shift invariant system, with impulse response $h(t)$ (Assume that the impulse response is real). We deal with stable LSI systems because the output will be bounded.

The output for an LSI system is given by the convolution of the input and the impulse response. The direct solution via convolution seems very difficult. In fact if we try to calculate the convolution integral, we get the following expression:

$$y(t) = \int_{-\infty}^{+\infty} h(\lambda) 2A\cos(\Omega(t - \lambda) + \phi) d\lambda$$

We will now see how to solve it using phasors, making the analysis mathematically convenient.

4. Sinusoids and LSI Systems

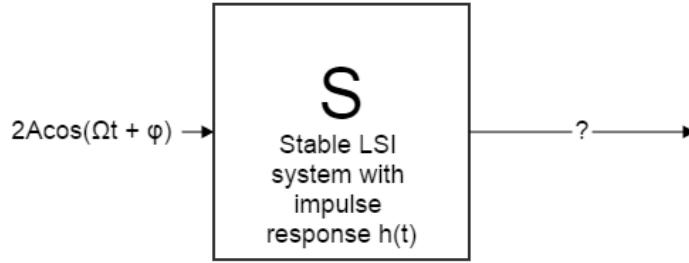


Figure 4.2.: Sinusoidal Input To Stable LSI System

4.2.2. Phasor Input to LSI System

Consider the output when a phasor is input to a stable LSI system.

$$\begin{aligned} Ae^{j(\Omega t + \phi)} &\xrightarrow{\text{LSI}} \int_{-\infty}^{\infty} Ae^{j(\Omega(t - \lambda) + \phi)} h(\lambda) d\lambda \\ &= Ae^{j(\Omega t + \phi)} \int_{-\infty}^{\infty} e^{-j\Omega\lambda} h(\lambda) d\lambda \end{aligned}$$

Here, we see that we get the output to be the input phasor with a multiplying factor $\int_{-\infty}^{\infty} e^{-j\Omega\lambda} h(\lambda) d\lambda$, which is dependent only on Ω and h (it is constant with respect to time).

We will now come back to the importance of why we chose our system to be stable. We will do this by trying to find a bound to the absolute value of the multiplying factor we just found.

$$\begin{aligned} \left| \int_{-\infty}^{\infty} e^{-j\Omega\lambda} h(\lambda) d\lambda \right| &\leq \int_{-\infty}^{\infty} |e^{-j\Omega\lambda}| |h(\lambda)| d\lambda \\ &= \int_{-\infty}^{\infty} |h(\lambda)| d\lambda \end{aligned}$$

And as per the condition of stability the integral of the absolute value of the impulse response is bounded. Hence the absolute value of the multiplying factor is also bounded. In fact we have obtained a concrete bound to this integral, namely the absolute integral of the impulse response. We can do similar mathematical analysis for the conjugate as well.

This multiplying factor is called the *Frequency Response* of the LSI system at angular frequency Ω .

$$\text{Frequency Response } H(\Omega) = \int_{-\infty}^{\infty} e^{-j\Omega\lambda} h(\lambda) d\lambda$$

4.2.3. Sinusoid Input to LSI System

We will now come back to our original problem of inputting a sinusoidal signal into a stable linear shift invariant system. We will begin by simplifying the outputs of the two phasors.

$$H(\Omega) Ae^{j(\Omega t + \phi)} = |H(\Omega)| A e^{j(\Omega t + \phi + \angle H(\Omega))}$$

The complex conjugate of the frequency response is given by

$$\bar{H}(\Omega) = \int_{-\infty}^{\infty} e^{j\Omega\lambda} h(\lambda) d\lambda = H(-\Omega)$$

Note that the first equality holds only if $h(\lambda)$ is real. Hence,

$$H(-\Omega)Ae^{-j(\Omega t+\phi)} = |H(\Omega)|Ae^{-j(\Omega t+\phi+\angle H(\Omega))}$$

So we can check by adding the two outputs that the output for inputting a sinusoid $2A\cos(\Omega t + \phi)$ is given by:

$$2A\cos(\Omega t + \phi) \xrightarrow{\text{Stable LSI}} 2A|H(\Omega)|\cos(\Omega t + \phi + \angle H(\Omega))$$

This indicates that on passing through a stable LSI system, a sinusoid goes through an amplitude and phase shift which is independent of time and only dependent on the impulse response and angular frequency of the sinusoid. Note again that the impulse response has to be real for this to hold true.

Also, if we view an input to a stable LSI system as the sum of multiple sinusoids, the output can be visualized as the sum of the corresponding outputs of the sinusoids through the stable LSI system. The stability of the LSI systems guarantees the existence of the frequency response of the system. But if it is unstable it may or may not have a frequency response.

4.2.4. Periodic Input to Shift Invariant Systems

How does an LSI system respond to a general periodic signal? Consider the following lemma.

Lemma: A periodic input to a shift invariant system produces a periodic output.

Proof:

Let the input be $x(t)$ with a period T

$$x(t+T) = x(t) \quad \forall t$$

$$x(t) \xrightarrow{\text{SI system}} y(t)$$

Now, by shift invariance,

$$x(t+T) \xrightarrow{\text{SI system}} y(t+T)$$

Hence,

$$y(t) = y(t+T) \quad \forall t \tag{4.1}$$

Hence the output is also periodic, infact with the same period T .

An important property of a periodic input or a periodic function is that it can be represented by a linear combination (which can be countably infinite) of sinusoids having frequencies which are multiples of the frequency of the input. This property is referred to as the existence of a Fourier Series expansion of the input.

The Fourier Series expansion and our understanding from the previous section gives us a simple method to compute the output of a periodic input in an LSI system as the linear combination of the output of these sinusoids.

This result, combined with 4.1 can be used to calculate the output for a general periodic input, but for that, we need still more information. How do we know whether a certain periodic signal can be expressed as a linear combination of sinusoids *uniquely*? To answer this question, we need to invoke some concepts from linear algebra by considering signals as vectors.

4.3. Module 2: Lecture 3

Signals and Vectors

4.3.1. Relations between Signals and Vectors

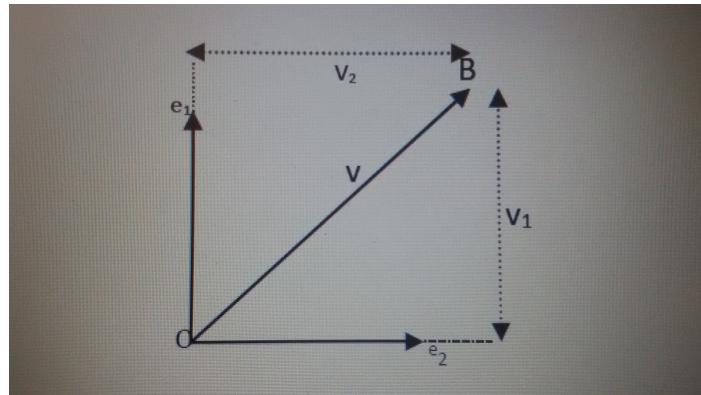
Consider a 2-dimensional vector. We can decompose this vector into two perpendicular components with e_1 and e_2 as the unit vectors along those directions. To find these components, we need to use the dot product.

Dot Product

The dot product of two vectors u and v is defined as the magnitude of the vector u multiplied by the magnitude of the vector v multiplied by the cosine of the angle between these two vectors.

$$u \cdot v = |u||v|\cos\theta$$

where θ is the angle between the vectors u and v . If u is a unit vector, then the dot product of vectors v and u gives the component of the vector v in the direction of u , with the value $|v|\cos\theta$. In Fig.??, OB is a vector denoted by v which is expressed as



$$v = v_1e_1 + v_2e_2$$

where v_1 and v_2 are the components in the e_1 and e_2 directions respectively. In this way, we can express any vector as a linear combination of e_1 and e_2 . Thus, we can say that the vectors e_1 and e_2 span the 2D space. Suppose if any vector can be expressed uniquely as a linear combination of $\{u_1, u_2, u_3, \dots, u_n\}$, where $\{u_1, u_2, u_3, \dots, u_n\}$ are linearly independent, then $\{u_1, u_2, \dots, u_n\}$ is said to form a basis.

Also, if $\langle u_i, u_j \rangle = 0$ for $i, j \in [1, n]$ and $i \neq j$, i.e. the dot product of any two distinct basis vectors is zero, then the basis is said to be an *orthogonal basis*. Finally, for a n -dimensional space, a collection of n linearly independent vectors forms the basis.

The dot product of two vectors u and v is equal to the sum of the products of the corresponding perpendicular components. Suppose

$$u = u_1e_1 + u_2e_2$$

$$v = v_1e_1 + v_2e_2$$

then,

$$u \cdot v = u_1v_1 + u_2v_2$$

Dot Product of Discrete Sequences

Now, consider a discrete sequence with 2 non-zero points. This sequence can be compared to a 2-dimensional vector v with v_1 and v_2 as its perpendicular components such that it is equal to the values at the 2 non-zero points of the sequence. Similarly for a sequence with n non-zero points can be considered as a vector in n -dimensional space. Also, the concept of dot product is similarly applied to the discrete sequences.

For example, consider

$$x[n] = (1/2)^n u[n]$$

i.e.

$$x[n] = \begin{cases} (1/2)^n & \text{for } n \geq 0 \\ 0 & \text{for } n < 0 \end{cases}$$

$$y[n] = (1/3)^n u[n]$$

i.e.

$$y[n] = \begin{cases} (1/3)^n & \text{for } n \geq 0 \\ 0 & \text{for } n < 0 \end{cases}$$

Let's calculate the dot product of $x[n]$ and $y[n]$, i.e. sum the product of the corresponding components. We have,

$$\begin{aligned} \langle x[n], y[n] \rangle &= \sum_{n=-\infty}^{\infty} x[n]y[n] \\ \langle x[n], y[n] \rangle &= \sum_{n=0}^{\infty} x[n]y[n] \\ \langle x[n], y[n] \rangle &= \sum_{n=0}^{\infty} (1/2)^n * (1/3)^n \\ \langle x[n], y[n] \rangle &= \sum_{n=0}^{\infty} (1/6)^n \\ \langle x[n], y[n] \rangle &= \frac{1}{(1 - (1/6))} \\ \langle x[n], y[n] \rangle &= \frac{6}{5} \end{aligned}$$

This is the dot product of the two discrete sequences. This dot product is also called as the *inner product*. The inner product of two vectors u and v is represented by $\langle u, v \rangle$.

Inner Product of Continuous Signals

In a discrete sequence, a unit vector can be expressed as $\delta[n - N]$ for all $N \in \mathbb{Z}$. An n -dimensional vector v can be expressed as

$$v = v_1 e_1 + v_2 e_2 + \dots + v_n e_n$$

Now writing the above equation in terms of sequences, we have

4. Sinusoids and LSI Systems

$$x[n] = \sum_{n=-\infty}^{\infty} x[N] \delta[n - N]$$

where $x[N]$ is the component along the dimension N .

Applying the same concept for continuous time functions, we need to replace the summation by an integral, which gives,

$$x(t) = \int_{-\infty}^{\infty} x(\lambda) \delta[t - \lambda] d\lambda$$

where $x(\lambda)$ is the component in the direction λ and $\delta[t - \lambda]$ is continuous impulse at λ similar to a unit vector. So, the inner product of $x(t)$ and $y(t)$ is given by

$$\langle x(t), y(t) \rangle = \int_{-\infty}^{\infty} x(t)y(t) dt$$

4.3.2. Sinusoids as basis vectors

Let us see how we can express any periodic signal as a linear combination of sinusoids. We first need to find sinusoidal signals which are perpendicular. Consider a signal $x(t)$ which is periodic with period T , i.e.

$$x(t) = x(t + T) \quad \forall t$$

Assume that it can be expressed as a sum of sinusoidal functions. We should take a sinusoidal function which is periodic with period T . So it is of the form $A_k \cos(\frac{2\pi}{T}kt + \phi_k)$. We have,

$$x(t) = \sum_{k=-\infty}^{\infty} A_k \cos\left(\frac{2\pi}{T}kt + \phi_k\right)$$

Let's take two different k 's.

$$x_1(t) = A_1 \cos\left(\frac{2\pi}{T}k_1 t + \phi_1\right)$$

and

$$x_2(t) = A_2 \cos\left(\frac{2\pi}{T}k_2 t + \phi_2\right)$$

Consider the inner product of $x_1(t)$ and $x_2(t)$ with the interval going from 0 to T . We are here restricting our interval to T because the integral might diverge when we integrate from 0 to ∞ . We have,

$$\langle x_1(t), x_2(t) \rangle = \int_0^T x_1(t)x_2(t) dt$$

$$\langle x_1(t), x_2(t) \rangle = \int_0^T \cos\left(\frac{2\pi}{T}k_1 t + \phi_1\right) \cos\left(\frac{2\pi}{T}k_2 t + \phi_2\right) dt$$

Using

$$2 \cos A \cos B = \cos\left(\frac{A+B}{2}\right) + \cos\left(\frac{A-B}{2}\right)$$

We have,

$$\langle x_1(t), x_2(t) \rangle = \int_0^T \frac{1}{2} \left\{ \cos\left(\frac{2\pi}{T} \frac{(k_1+k_2)}{2} t + \frac{\phi_1+\phi_2}{2}\right) + \cos\left(\frac{2\pi}{T} \frac{(k_1-k_2)}{2} t + \frac{\phi_1-\phi_2}{2}\right) \right\} dt$$

If $(k_1 + k_2)$ or $(k_1 - k_2)$ are not zero, that means we have a finite number of cycles of the sinusoids, implying the integral is zero. However if $k_1 = k_2$, the second integral becomes $T/2$ times $\cos((\phi_1 - \phi_2)/2)$. Thus we have,

$$\langle x_1(t), x_2(t) \rangle \neq 0 \quad \text{for } k_1 = k_2$$

$$\langle x_1(t), x_2(t) \rangle = 0 \quad \text{for } k_1 \neq k_2$$

Hence, we have proved that two sinusoids with the same time period are perpendicular if they don't have the same angular frequency and vice-versa. Using this important concept, we will be able to write any periodic signal as sum of sinusoidal signals and hence analysis of these signals will be simpler.

4.4. Module 2: Lecture 4

Decomposition of Signals

4.4.1. Introduction

A fundamental idea in the study of signals is to represent signals in terms of linear combination of *basis* signals. In the previous section, we have embarked on some important concepts, relating to the idea of representing signals as a linear combination of sinusoids. We assumed that we wish to look either at periodic signals or signals over a finite interval T . Without loss of generality, let that interval be $(0, T)$.

We have assumed that such signals are signals spanned by all sinusoids of angular frequency $\Omega = \frac{2\pi}{T}kt, k = 0, \pm 1, \pm 2..$

The typical periodic function, with period T is of the form

$$x(t) = \sum_{k=-\infty}^{\infty} A_k \cos\left(\frac{2\pi}{T}kt + \phi_k\right)$$

where A_k denotes the amplitude, $\Omega = \frac{2\pi}{T}k$ is the angular frequency and ϕ_k the phase. We had proved that these sinusoids are orthogonal if they don't have same angular frequency.

$$\langle A_{k_1} \cos\left(\frac{2\pi}{T}k_1t + \phi_1\right), A_{k_2} \cos\left(\frac{2\pi}{T}k_2t + \phi_2\right) \rangle = 0 \quad \text{for } k_1 \neq k_2$$

This important concept requires appreciating several different ideas. Using these ideas, we can decompose signal $x(t)$ in terms of sinusoids. However, there are some signals which can't be spanned using these sinusoids. Such signals are beyond the scope of discussion that we are going to undertake at this stage.

4.4.2. Orthogonal vectors at a particular frequency

Consider a periodic signal $x(t)$ with fundamental period T . Then we can write $x(t)$ as

$$x(t) = \sum_{k=-\infty}^{\infty} A_k \cos\left(\frac{2\pi}{T}kt + \phi_k\right)$$

We constrain ourselves to $0 < t < T$.

To calculate A_k and ϕ_k we need to use the following property i.e.

$$\langle \cos\left(\frac{2\pi}{T}k_1t + \phi_1\right), \cos\left(\frac{2\pi}{T}k_2t + \phi_2\right) \rangle = 0 \quad \text{for } k_1 \neq k_2$$

4. Sinusoids and LSI Systems

We can think of the cosines $\cos(\frac{2\pi}{T}kt + \phi_k)$ as countably infinite orthogonal vectors, indexed by k , for $k = 0, \pm 1, \pm 2, \dots$ and so on.

To find out the component of $x(t)$ along these orthogonal vectors, we need to take the dot product of $x(t)$ with the unit vector, in the direction of orthogonal vectors $\cos(\frac{2\pi}{T}kt + \phi_k)$, $0 < t < T$. For calculating unit vector, we need to find the magnitude of the vector $\cos(\frac{2\pi}{T}kt + \phi_k)$ using dot product.

$$\begin{aligned} \left| \cos\left(\frac{2\pi}{T}kt + \phi_k\right) \right| &= \langle \cos\left(\frac{2\pi}{T}kt + \phi_k\right), \cos\left(\frac{2\pi}{T}kt + \phi_k\right) \rangle \\ &= \int_0^T \cos\left(\frac{2\pi}{T}kt + \phi_k\right) \cos\left(\frac{2\pi}{T}kt + \phi_k\right) dt \\ &= \int_0^T \cos^2\left(\frac{2\pi}{T}kt + \phi_k\right) dt \\ &= \int_0^T \left(\frac{1}{2} + \cos(2(\frac{2\pi}{T}kt + \phi_k))\right) dt \\ &= \int_0^T \frac{1}{2} dt + \int_0^T \cos(2(\frac{2\pi}{T}kt + \phi_k)) dt \\ &= \frac{T}{2} \end{aligned}$$

Thus, the magnitude square of the cosine is $T/2$ and hence the magnitude is $\sqrt{T/2}$. So, the unit vector is given by

$$\text{Unit vector} = \sqrt{\frac{2}{T}} \cos\left(\frac{2\pi}{T}kt + \phi_k\right)$$

But still, there is one unknown quantity, namely, ϕ_k in this unit vector. However, we can think of this vector as a vector in a two dimensional space:

$$\sqrt{\frac{2}{T}} \cos\left(\frac{2\pi}{T}kt + \phi_k\right) = \sqrt{\frac{2}{T}} [\cos \phi_k \cos\left(\frac{2\pi}{T}kt\right) - \sin \phi_k \sin\left(\frac{2\pi}{T}kt\right)]$$

that is, the unit vector can be thought of as a linear combination of two vectors.

We have,

$$\begin{aligned} \langle \cos\left(\frac{2\pi}{T}kt + \phi_k\right), \sin\left(\frac{2\pi}{T}kt + \phi_k\right) \rangle &= \int_0^T \cos\left(\frac{2\pi}{T}kt + \phi_k\right) \sin\left(\frac{2\pi}{T}kt + \phi_k\right) dt \\ &= \frac{1}{2} \int_0^T 2 \cos\left(\frac{2\pi}{T}kt + \phi_k\right) \sin\left(\frac{2\pi}{T}kt + \phi_k\right) dt \\ &= \frac{1}{2} \int_0^T \sin\left(\frac{4\pi}{T}kt + \phi_k\right) dt \\ &= 0 \end{aligned}$$

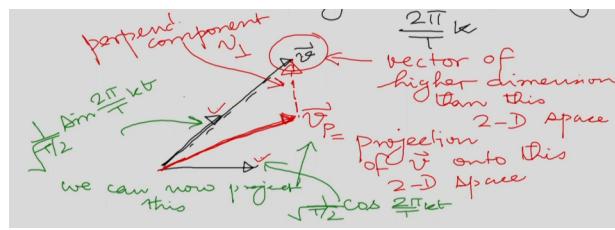
Hence, the cosine and sine components are also orthogonal. $\cos \phi_k$ and $\sin \phi_k$ are now two unknown constants. Therefore, decomposing $x(t)$ along $\cos(\frac{2\pi}{T}kt + \phi_k)$ will give the same results as decomposing $x(t)$ along $\cos(\frac{2\pi}{T}kt)$ and $\sin(\frac{2\pi}{T}kt)$. So, to decompose $x(t)$ along $\cos(\frac{2\pi}{T}kt)$ and $\sin(\frac{2\pi}{T}kt)$, we require unit vectors, which we already have calculated above. So, the two unit vectors are $\sqrt{\frac{2}{T}} \cos\left(\frac{2\pi}{T}kt\right)$ and $\sqrt{\frac{2}{T}} \sin\left(\frac{2\pi}{T}kt\right)$.

4.4.3. Projection of Vectors

Let $u_1 = \sqrt{\frac{2}{T}} \cos(\frac{2\pi}{T} kt)$ and $u_2 = \sqrt{\frac{2}{T}} \sin(\frac{2\pi}{T} kt)$. We know that at frequency $2\pi k/T$, there are two orthogonal vectors, u_1 and u_2 . We have a two dimensional space and we need to calculate the component of the vector $x(t)$ along orthogonal vectors of this 2D space.

Let v be any arbitrary vector of higher dimension and v_p be the projection of v onto our 2D space spanned by orthogonal vectors u_1 and u_2 . To calculate the components of the vector v along the two orthogonal vectors, we can project v_p onto u_1 and u_2 . We can also directly project v onto u_1 and u_2 .

The vector v can be written as $v = v_p + v_{\perp}$, where v_p is the projection of v on to 2D space and v_{\perp} is component perpendicular to the 2D space. The projection of v onto vectors u_1 and u_2 will be same as the projection of v_p on those vectors, as the v_{\perp} will not contribute in projection.



$$\begin{aligned} & \text{2-D space at this angular frequency } \frac{2\pi k}{T} \\ & \frac{1}{\sqrt{2}} \cos\left(\frac{2\pi k t}{T}\right) = \hat{u}_2 \\ & \frac{1}{\sqrt{2}} \cos\left(\frac{2\pi k t}{T}\right) = \hat{u}_1 \\ & \langle \vec{v}_p, \hat{u}_1 \rangle = v_{p1} \\ & \langle \vec{v}_p, \hat{u}_2 \rangle = v_{p2} \end{aligned}$$

Let us project v_p along the two orthogonal vectors. For projecting v_p along u_1 and u_2 , taking dot product

$$\begin{aligned} \langle v_p, u_1 \rangle &= v_{p1} \\ \langle v_p, u_2 \rangle &= v_{p2} \end{aligned}$$

v_p in terms of u_1, u_2 can be written as $v_p = v_{p1}u_1 + v_{p2}u_2$. Interestingly $\langle v_p, u_1 \rangle = \langle v, u_1 \rangle$ is true also for u_2 as the perpendicular component v_{\perp} is not going to contribute in the dot product.

We know that the k^{th} angular frequency = $2\pi k/T$. At this frequency we have two components

$$x(t) = \langle x(t), \sqrt{\frac{2}{T}} \cos\left(\frac{2\pi}{T} kt\right) \rangle \cdot \sqrt{\frac{2}{T}} \cos\left(\frac{2\pi}{T} kt\right) + \langle x(t), \sqrt{\frac{2}{T}} \sin\left(\frac{2\pi}{T} kt\right) \rangle \cdot \sqrt{\frac{2}{T}} \sin\left(\frac{2\pi}{T} kt\right)$$

This can be simplified as

$$x(t) = \langle x(t), \cos\left(\frac{2\pi}{T} kt\right) \rangle \cdot \frac{2}{T} \cos\left(\frac{2\pi}{T} kt\right) + \langle x(t), \sin\left(\frac{2\pi}{T} kt\right) \rangle \cdot \frac{2}{T} \sin\left(\frac{2\pi}{T} kt\right)$$

4. Sinusoids and LSI Systems

4.4.4. Decomposition of square wave

We have a periodic signal $x(t)$ in Fig.??, defined as

$$x(t) = \begin{cases} 1 & \text{if } 0 < t < \frac{T}{2} \\ -1 & \text{if } \frac{T}{2} < t < T \end{cases}$$

$x(t) = x(t + T)$ for all t , i.e. $x(t)$ is periodic with period T .

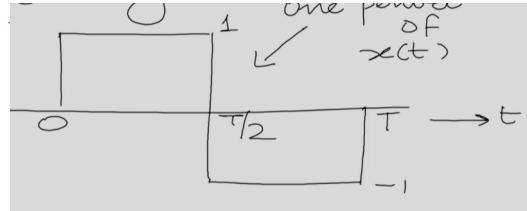


Figure 4.3.: Plot of $x(t)$

The zero frequency component is the mean of $x(t)$ over a period T which is 0 since the net area under $x(t)$ in one period is zero.

Let us calculate the quantity $\langle x(t), \cos(\frac{2\pi}{T}kt) \rangle \cdot \frac{2}{T} \cos(\frac{2\pi}{T}kt)$,

For $k = 1$ and $k = 2$, we can see from Fig.?? that, the positive portion gets annulled by the negative portion. Similar is the case for $k = 3, 4, \dots$ and so on. Hence $\langle x(t), \cos(\frac{2\pi}{T}kt) \rangle = 0$ for given $x(t)$.

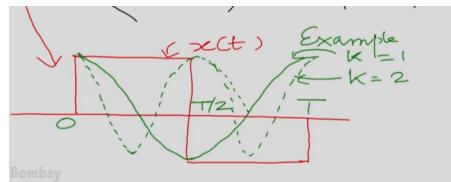


Figure 4.4.: Dot product of the square wave with cosines

Calculating $\langle x(t), \sin(\frac{2\pi}{T}kt) \rangle \cdot \frac{2}{T} \sin(\frac{2\pi}{T}kt)$, we have as shown in Fig.??,

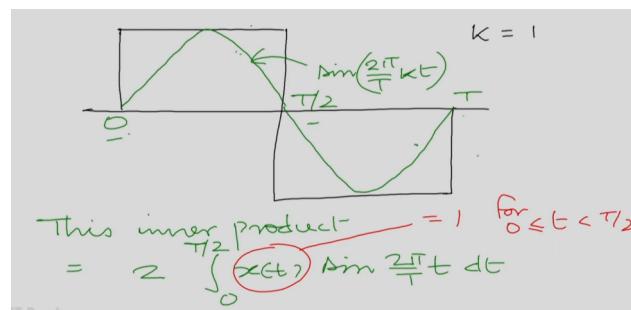


Figure 4.5.: Dot product of the square wave with sines

$$\langle x(t), \sin(\frac{2\pi}{T}kt) \rangle = \int_0^T x(t) \sin(\frac{2\pi}{T}kt) dt$$

$$\langle x(t), \sin(\frac{2\pi}{T}kt) \rangle = \int_0^{\frac{T}{2}} \sin(\frac{2\pi}{T}kt) dt + \int_{\frac{T}{2}}^T -\sin(\frac{2\pi}{T}kt) dt$$

Also, we know that

$$\int_0^{\frac{T}{2}} \sin\left(\frac{2\pi}{T}kt\right) dt = - \int_{\frac{T}{2}}^T \sin\left(\frac{2\pi}{T}kt\right) dt$$

Using this, we have

$$\langle x(t), \sin\left(\frac{2\pi}{T}kt\right) \rangle = 2 \int_0^{\frac{T}{2}} \sin\left(\frac{2\pi}{T}kt\right) dt = \left[-2 \frac{\cos\left(\frac{2\pi}{T}kt\right)}{\frac{2\pi}{T}k} \right]_0^{\frac{T}{2}}$$

For $k = 2n$, the above integral goes to zero as $\cos(2\pi n) = 1$ for all $n \in \mathbb{Z}$. For $k = 2n-1$, the above integral has the value of $4T/\pi k$ because $\cos(2\pi(2n-1)) = -1$ for all $n \in \mathbb{Z}$. Thus,

$$\langle x(t), \sin\left(\frac{2\pi}{T}kt\right) \rangle = \frac{4T}{\pi k} \text{ (for odd } k)$$

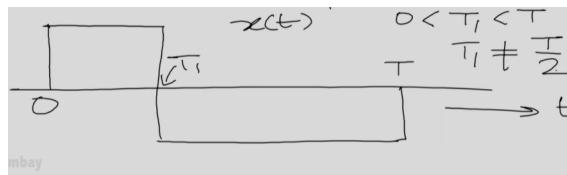
Hence, we have, after multiplying by $2/T$,

$$x(t) = \sum_{n=1}^{\infty} \frac{4}{\pi(2k-1)} \sin\left(\frac{2\pi}{T}(2k-1)t\right)$$

4.4.5. Examples of Decomposition

Exercise 1

Work out the decomposition for an asymmetric square wave, of period T as shown in the figure below.



$$x(t) = 1 \quad 0 < t < T_1$$

$$x(t) = -1 \quad T_1 < t < T$$

$$0 < T_1 < T, \quad T_1 \neq T/2, \quad x(t+T) = x(t) \quad \forall t \in R.$$

Exercise 2

In the answer of exercise 1 put $T_1 = T/2$, and verify the answer with that of decomposition of symmetric square wave.

Exercise 3

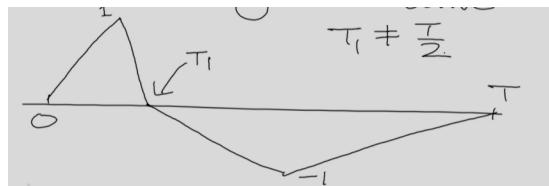
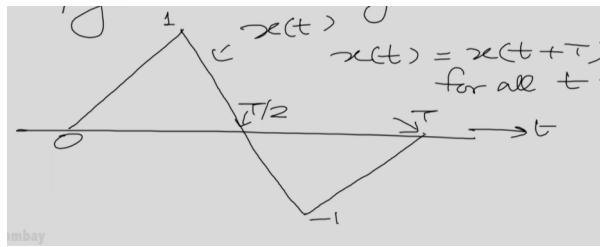
Work out the decomposition for symmetric triangular wave, periodic with period T .

Hint: Use and observe the symmetry graphically.

Exercise 4

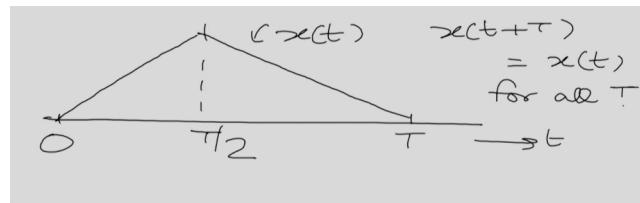
Evaluate the decomposition for an asymmetric triangular wave as shown in the figure. Cross check the answer with that of exercise 3, by putting $T_1 = T/2$ in the final answer.

4. Sinusoids and LSI Systems



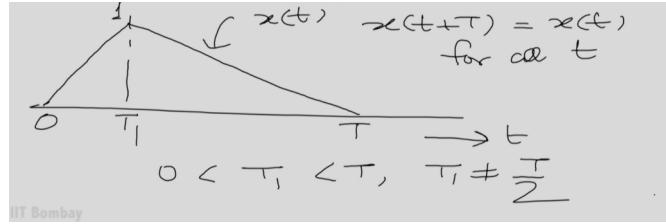
Exercise 5

A variant of symmetric triangular wave is shown below. Find its decomposition and compare with answer of exercise 3.



Exercise 6

A variant of an asymmetric triangular wave is shown below. Evaluate the decomposition of this wave. Substitute $T_1 = T/2$, in the answer and compare with the result of exercise 5. Explain the differences in the result, if any.



4.5. Module 2: Lecture 5

Fourier Series Decomposition and its Applications

4.5.1. Introduction

We have been decomposing periodic waveforms into their periodic sinusoidal components. In this section, we will look at a variation of the same decomposition using complex exponential functions. One way of doing this is to convert each sinusoidal component of sinusoidal decomposition into corresponding exponential functions. However, in some cases, we prefer to *a priori* decompose the waveform in complex exponential components. We shall also see how to go from the complex exponential components to sinusoidal components. We will then look at the applications of both sinusoidal and complex exponential decompositions; in particular, we will be analyzing its utility in determining the output from a linear shift invariant system on applying a periodic input. We will also see some conditions under which we cannot obtain a Fourier series decomposition for a periodic waveform.

4.5.2. Complex Exponential Fourier Series Decomposition

Suppose we are given a periodic signal $x(t)$ with period T . We wish to write $x(t)$ in terms of sum of complex exponential functions:

$$x(t) = \sum_{k=-\infty}^{\infty} C_k e^{j2\pi kt/T}$$

where $C_k \in \mathbb{C}$. Basically, we are trying to write $x(t)$ as a linear combination of complex exponential functions rotating at angular frequencies $2\pi k/T$ for all integer k . Note that the complex exponentials with angular frequencies as positive integral multiples of $2\pi k/T$ represent anticlockwise rotating phasors while the complex exponentials with negative angular frequencies represent clockwise rotating phasors.

Orthogonality of Fourier Series Components

We now establish the orthogonality between the complex exponential components. We take two integers k and l and evaluate the inner product

$$\int_{-T/2}^{T/2} \frac{1}{T} e^{(j2\pi kt/T)} e^{(j2\pi(-l)t/T)} dt = \int_{-T/2}^{T/2} \frac{1}{T} e^{(j2\pi(k-l)t/T)} dt \quad (4.2)$$

4. Sinusoids and LSI Systems

We have two cases now:

- **Case 1:** $k = l$: The integral in 4.2 becomes $\int_{-T/2}^{T/2} \frac{1}{T} dt = 1$
- **Case 2:** $k \neq l$: The integral in 4.2 becomes

$$\int_{-T/2}^{T/2} \frac{1}{T} e^{(j2\pi(k-l)t/T)} dt = \frac{1}{T} \frac{e^{(j2\pi(k-l)t/T)}}{(j2\pi(k-l)/T)} \Big|_{-T/2}^{T/2} = 0$$

Obtaining Complex exponential Decomposition from Sinusoidal Decomposition

Suppose we are given the sinusoidal components of the decomposition of a real waveform

$$x(t) = \sum_0^{\infty} A_k \cos(2\pi kt/T + \phi_k)$$

Note that in the above sum, for $k = 0$, we will have the component as $A_0 \cos(\phi_0)$. We can merge the constant $\cos(\phi_0)$ into A_0 and we will simply have the constant A_0 for $k = 0$. Now for a non-zero integer k using $\cos(\theta) = \frac{1}{2}(e^{j\theta} + (e^{j\theta})^*)$, (Note: C^* denotes complex conjugate of C), we have

$$A_k \cos(2\pi kt/T + \phi_k) = \frac{1}{2} \left(A_k e^{j\phi_k} e^{j2\pi kt/T} + (A_k e^{j\phi_k})^* (e^{-j2\pi kt/T}) \right)$$

since A_k are real.

$$\begin{aligned} \implies C_k &= \frac{1}{2} A_k e^{j\phi_k} \\ \implies C_{-k} &= \frac{1}{2} (A_k e^{j\phi_k})^* \end{aligned}$$

And we also have trivially $C_0 = A_0$. Notice that for real waveform, we have $C_k = (C_{-k})^*$.

Converting Complex Exponential to Sinusoidal Decomposition

Suppose we are given a real signal

$$x(t) = \sum_{k=1}^{\infty} C_k e^{(j2\pi kt/T)} + \sum_{k=-\infty}^{-1} C_k e^{(j2\pi kt/T)} + C_0$$

Now, note that for an integer $k \neq 0$,

$$\int_{-T/2}^{T/2} e^{(j2\pi kt/T)} dt = \frac{1}{T} \frac{e^{(j2\pi kt/T)}}{(j2\pi k/T)} \Big|_{-T/2}^{T/2} = 0$$

whereas, for $k = 0$,

$$\int_{-T/2}^{T/2} 1 dt = T$$

So, basically we can obtain C_0 by integrating $x(t)$ over time T . In the Fourier decomposition, only C_0 will give a non-zero integral which is equal to $C_0 \cdot T$. Hence,

$$C_0 = \frac{1}{T} \int_{-T/2}^{T/2} x(t) dt$$

Now, since for real $x(t)$ we have $C_k = (C_{-k})^*$, we get

$$C_k e^{j2\pi kt/T} = (C_{-k} e^{-j2\pi kt/T})^*$$

So we have

$$C_k e^{j2\pi kt/T} + C_{-k} e^{-j2\pi kt/T} = 2\operatorname{Re}(C_k e^{j2\pi kt/T})$$

Using the polar form of C_k which is $|C_k|e_k^\phi$, we get

$$2\operatorname{Re}(C_k e^{j2\pi kt/T}) = 2|C_k| \operatorname{Re}(e^{j2\pi kt/T + \phi_k}) = 2|C_k| \cos(j2\pi kt/T + \phi_k)$$

Finally, we have the sinusoidal decomposition

$$x(t) = C_0 + \sum_{k=1}^{\infty} 2|C_k| \cos(j2\pi kt/T + \phi_k)$$

4.5.3. Periodic Input to a Simple RC Circuit

Suppose, we apply a real periodic voltage $x(t)$ with a time period T as an input to a series RC-circuit. Recall that an RC circuit is a linear shift invariant system. Suppose we can write complex exponential decomposition of $x(t)$ as

$$\sum_{k=-\infty}^{\infty} C_k e^{(j2\pi kt/T)}$$

The advantage of doing this is that we can use the fact that a complex exponential input to a linear shift invariant system simply gives the same complex exponential multiplied by a constant as its output. By phasor analysis of the circuit, we have the transfer function

$$\frac{j(2\pi k/T)CR}{1 + j(2\pi k/T)CR}$$

So, the output waveform will be simply

$$y(t) = \sum_{k=-\infty}^{\infty} \frac{j(2\pi k/T)CR}{1 + j(2\pi k/T)CR} C_k e^{(j2\pi kt/T)}$$

This illustrates the power of having complex exponential decomposition of a waveform since we can obtain the output waveform quite simply if it is passed through a linear shift invariant system. Now, one catch in the above discussion is that we assumed that the decomposition of $x(t)$ into a Fourier series is possible. It turns out it's *not* always the case that we can write a Fourier series decomposition for a periodic waveform! Although, for most of the practical waveforms, we can write it. We will look at the conditions under which we can write the Fourier series decomposition for a periodic waveform. There are some waveforms for which we cannot find the Fourier decomposition. One such example is $x(t) = \sin(1/\operatorname{frac}(t))$ where $\operatorname{frac}(t)$ denotes fractional part of t . So, $x(t)$ is periodic with period 1 but still its Fourier series decomposition fails to exist. There are certain conditions called the 'Dirichlet' conditions under which Fourier analysis can be done. We shall not discuss them here.

4. Sinusoids and LSI Systems

4.5.4. Periodic Inputs to a General Linear Shift Invariant System

In this section we will use the Fourier series decomposition to determine the output from a linear shift invariant system on applying a periodic input. Of course, for our analysis, we will assume that the Fourier series decomposition of the input waveform exists.

Let \mathbb{S} denote a linear shift invariant system and let $h(t)$ be its impulse response. We apply a periodic input waveform $x(t)$. Now, using linearity of the system, output of the system is simply the sum of the outputs obtained for each component of Fourier series decomposition of $x(t)$.

Let us derive the output for the k^{th} component of $x(t)$. The output for k^{th} will be the convolution of $h(t)$ and the input itself.

$$\int_{-\infty}^{\infty} h(\tau) \cdot C_k e^{j2\pi k(t-\tau)/T} d\tau$$

Since the above integral runs over τ and not t , the expression reduces to

$$C_k e^{j2\pi kt/T} \int_{-\infty}^{\infty} h(\tau) \cdot e^{-j2\pi k\tau/T} d\tau$$

The integral reduces to a constant dependant on k . So, what we obtain is quite interesting as it is simply the input itself multiplied by a constant! Let us denote the constant by $\mathcal{H}(k)$. So, the output, which is the sum of the outputs obtained for each component, is

$$\sum_{k=-\infty}^{\infty} \mathcal{H}(k) C_k e^{j2\pi kt/T}$$

4.6. Conclusion

In this chapter, we discussed about sinusoidal inputs to LSI systems, Fourier series decomposition, its properties and its application in analysing linear shift invariant systems. In the coming chapters, we will discuss the significance of the quantity $\mathcal{H}(k)$ we derived in previous section and begin with what is known as the Fourier Transform.

5. The Fourier Transform

5.1. Module 2: Lecture 6 The Fourier Transform

5.1.1. Introduction

We are now in a position where we can deal with more general signals from the point of view of decomposition into sinusoidal frequency components and analysis as to what happens when you pass them through a linear shift invariant system. To do that, let's recapitulate some of the important conclusions that we had drawn in the previous chapter which will now help us in generalization.

5.1.2. Recapitulation

Let \mathbb{S} be a linear shift invariant system with impulse response $h(t)$, and to it, we give a periodic signal input. Let's assume that the Dirichlet conditions are obeyed. So we could decompose this into its Fourier series.

$$x(t) = \sum_{k=-\infty}^{k=+\infty} c(k)e^{j\Omega kt}$$

Where $\Omega = 2\pi/T$ is the angular frequency and $c(k)$ are the Fourier coefficients. Now, the output can also be written as a Fourier series.

$$y(t) = \sum_{k=-\infty}^{k=+\infty} c(k)H(k\Omega)e^{j\Omega kt}$$

where

$$H(\Omega) = \int_{-\infty}^{+\infty} h(t)e^{-j\Omega t} dt$$

5.1.3. Interpretation of $H(\Omega)$ as a Dot Product

We can interpret $H(\Omega)$ as a dot product or inner product. It's an inner product between the impulse response $h(t)$ and the rotating phasor $e^{j\Omega t}$. An inner product with a unit vector calculates the projection or a component of the vector along the unit vector. So it's as if we are trying to find the component of the impulse response along a rotating phasor, rotating with an angular velocity Ω . And, at the specific values of Ω given by each of the Fourier series components, we evaluate this dot product and use it to modify the Fourier series coefficient. We can see from the equation of output $y(t)$ that output Fourier series coefficients are $c(k)H(\Omega k)$. Recall that in the input, the Fourier series coefficients were calculated by taking an inner product taking only one period of the input for the integral. We took only one period of the input, and found the inner product with the corresponding harmonic or the complex exponential rotating with that particular multiple of the fundamental frequency. This is how we calculate the Fourier series coefficients.

5. The Fourier Transform

Now, we can see that the Fourier coefficients of the output are the *component by component multiplication of the Fourier coefficients of the input and impulse response*.

$$Y(k\Omega) = c(k)H(k\Omega)$$

Now, there are two questions to answer here. Here we have assumed that the input was periodic. What if it is not? Can we generalize this to a non-periodic function is a question that we need to answer. Secondly, can we think of this quantity $H(\Omega)$ as a new transform or a new way of dealing with the impulse response in its own right?

Now, a transform will make sense or will be adequate only if it is invertible. That is to say, we should be able to get back $h(t)$ from $H(\Omega)$. There should be no loss of information. It turns out that this is indeed the case. We can call $H(\Omega)$, for any $\Omega \in \mathbb{R}$, as the *Fourier transform* of the impulse response $h(t)$.

Now let's try to address the question as to how can we *go back* to $h(t)$ from $H(\Omega)$.

5.1.4. Inverse for the Fourier Transform

The answer to the question again lies in vector intuition. As we have seen, $H(\Omega)$ is the component of $h(t)$ along the phasor $e^{j\Omega t}$ for all such $\Omega \in \mathbb{R}$. In vector algebra, we get back a vector from its components by multiplying the components with their corresponding unit vectors and summing them. So, if \hat{u}_1 and \hat{u}_2 are two orthogonal unit vectors in two dimensions, \vec{v} is a 2-D vector, and

$$\langle \vec{v}, \hat{u}_1 \rangle = v_1, \langle \vec{v}, \hat{u}_2 \rangle = v_2$$

Then, we construct \vec{v} by multiplying the components with the corresponding unit vector and summing them. Hence

$$\vec{v} = v_1 \hat{u}_1 + v_2 \hat{u}_2$$

Now using the same principle here, $H(\Omega)$ are the components of the “vector” $h(t)$ along the unit vectors $e^{j\Omega t}$. Hence, to obtain the vector $h(t)$ back, we should multiply the component by the unit vector and sum up, or in this case, integrate over all Ω .

But there are two subtleties involved here. Firstly, we don't know whether in this vector space of integrable functions, the vector $e^{j\Omega t}$ is a unit vector or not. That is, whether its magnitude is one or not. In case it is not one, we have to multiply it by a suitable constant to make it a unit vector. Secondly, we don't know whether $e^{j\Omega_1 t}$ and $e^{j\Omega_2 t}$ are orthogonal or not, for $\Omega_1 \neq \Omega_2$.

In the following subsection, we proceed to evaluate the inverse of $H(\Omega)$ assuming two things; one is that the vectors $e^{j\Omega t}$ are orthogonal, and that the multiplying factor κ_0 (for making it a unit vector) is independent of Ω .

Finding the Inverse of the Fourier Transform

From the previous discussion, the quantity we propose to be the inverse of $H(\Omega)$ is given by

$$I = \int_{-\infty}^{\infty} H(\Omega) \kappa_0 e^{j\Omega t} d\Omega$$

Now, we know that

$$H(\Omega) = \int_{-\infty}^{\infty} h(t_1) e^{-j\Omega t_1} dt_1$$

Putting this in I , we get,

$$\begin{aligned} I &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(t_1) e^{-j\Omega t_1} dt_1 \kappa_0 e^{j\Omega t} d\Omega \\ &= \int_{-\infty}^{\infty} h(t_1) \left\{ \int_{-\infty}^{\infty} \kappa_0 e^{-j\Omega t_1} e^{j\Omega t} d\Omega \right\} dt_1 \end{aligned}$$

Let's evaluate the quantity in the braces. We can write it as

$$\begin{aligned} &\lim_{\Omega_1 \rightarrow \infty} \int_{-\Omega_1}^{+\Omega_1} \kappa_0 e^{j\Omega(t-t_1)} d\Omega \\ &= \lim_{\Omega_1 \rightarrow \infty} \kappa_0 \frac{e^{j\Omega_1(t-t_1)} - e^{-j\Omega_1(t-t_1)}}{j(t-t_1)} \\ &= \lim_{\Omega_1 \rightarrow \infty} \frac{2\kappa_0 j \sin(\Omega_1(t-t_1))}{j(t-t_1)} \\ &= \lim_{\Omega_1 \rightarrow \infty} \frac{2\kappa_0 \Omega_1 \sin(\Omega_1(t-t_1))}{\Omega_1(t-t_1)} \end{aligned}$$

Now, this can be written as

$$\lim_{\Omega_1 \rightarrow \infty} 2\kappa_0 \Omega_1 \frac{\sin(x)}{x}$$

where $x = \Omega_1(t-t_1)$. Now, $\sin(x)/x$ is not defined at $x = 0$ as it has a zero divided by zero form. So we evaluate the limit as $x \rightarrow 0$.

$$\lim_{x \rightarrow 0} \frac{\sin(x)}{x} = \lim_{x \rightarrow 0} \frac{\cos(x)}{1} = 1$$

where the first equality is due to L'Hôpital's rule. This function will be oscillatory due to $\sin(x)$, but will decay as x increases, due to the x in the denominator. Also, the function will be zero when $x = n\pi$ or $\Omega_1(t-t_1) = n\pi$. Now, let's deal with this function as a function of $(t-t_1)$ rather than x . Let us try to plot the function

$$f(t-t_1) = \frac{2\kappa_0 \Omega_1 \sin(\Omega_1(t-t_1))}{\Omega_1(t-t_1)}$$

The value of $\sin(x)/x$ is 1 as $x \rightarrow 0$ as we saw earlier. Hence $f(0) = 2\kappa_0 \Omega_1$. Also the first null of the function will occur at $\Omega_1(t-t_1) = \pi$ or $(t-t_1) = \pi/\Omega_1$. Also the function is even with respect to $(t-t_1)$. Hence, we will get a plot roughly as shown in Fig.???. Now, we have to look at what happens when $\Omega_1 \rightarrow \infty$. We can see that the height of the main lobe at $x = 0$ goes to infinity, but the width of the lobe, which is $2\pi/\Omega_1$ goes to zero, as $\Omega_1 \rightarrow \infty$. It can also be shown that the area of the other smaller lobes add up to zero, due to their alternating nature. Hence, we have a *pulse* whose height is tending to infinity and the width to zero. In that case, we can assume the central lobe to be a triangle. Hence its area will be

$$A = \frac{1}{2} \frac{2\pi}{\Omega_1} 2\kappa_0 \Omega_1 = 2\pi\kappa_0$$

Hence, we can see that this area is a constant independent of Ω_1 . So we have a pulse having infinite magnitude and zero width, with a constant area encapsulated by it, and

5. The Fourier Transform

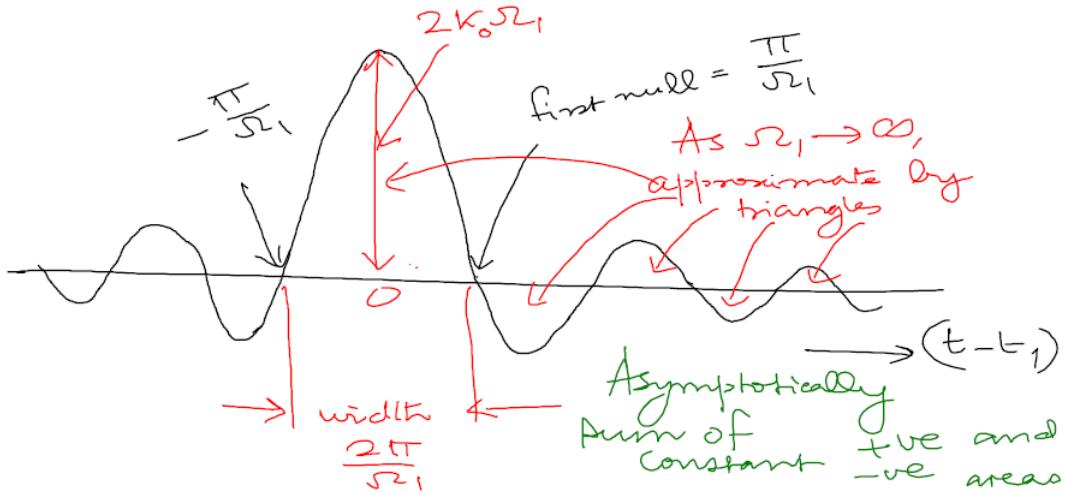


Figure 5.1.: Sinc function

as you will remember from Part 1, this is nothing but the *impulse function* $2\pi\kappa_0\delta(t - t_1)$. Now, going back to our integral I ,

$$I = \int_{-\infty}^{\infty} h(t_1) \left\{ \int_{-\infty}^{+\infty} \kappa_0 e^{-j\Omega t_1} e^{j\Omega t} d\Omega \right\} dt_1$$

And we have evaluated the quantity in the braces to be $2\pi\kappa_0\delta(t - t_1)$. Hence,

$$I = \int_{-\infty}^{\infty} h(t_1) \{2\pi\kappa_0\delta(t - t_1)\} dt_1 = 2\pi\kappa_0 \int_{-\infty}^{\infty} h(t_1) \delta(t - t_1) dt_1$$

Hence, by the sifting property of the delta function, we have,

$$I = 2\pi\kappa_0 h(t)$$

Hence, we can now set the value of κ_0 such that $I = h(t)$ as required. Hence $\kappa_0 = 1/2\pi$. Hence, we finally have our expression for the *Inverse Fourier transform* of $H(\Omega)$.

$$h(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} H(\Omega) e^{j\Omega t} d\Omega$$

5.1.5. Orthogonality of the Complex Exponentials

A valid question to ask here is where did the orthogonality of the complex exponentials come in the picture. The answer lies in the integral in the braces which we evaluated.

$$\int_{-\infty}^{+\infty} \kappa_0 e^{-j\Omega t_1} e^{j\Omega t} d\Omega = 2\pi\kappa_0\delta(t - t_1)$$

or,

$$\int_{-\infty}^{+\infty} e^{j\Omega t} \overline{e^{j\Omega t_1}} d\Omega = 2\pi\delta(t - t_1)$$

where the bar indicates complex conjugate. Now, this can be also written by renaming the variables, treating t as the independent variable, as

$$\int_{-\infty}^{+\infty} e^{j\Omega_1 t} \overline{e^{j\Omega_2 t}} dt = 2\pi\delta(\Omega_1 - \Omega_2)$$

5.2. Module 2: Lecture 7 Fourier Transforms properties and Fourier series example

Now, we can immediately see that this is like an inner product between phasors rotating with different frequencies Ω_1 and Ω_2 , and what this statement says is that it is non zero only for $\Omega_1 = \Omega_2$. Hence, we do have some sense of orthogonality here. But since the value is infinite when the frequencies are equal, we have to call this a generalized orthogonality relation.

5.1.6. Implications

Let's try to find what the Fourier transform expression implies. We have a non-periodic function $h(t)$ and we have decomposed it in its Fourier coefficients, not for discrete, but continuous set of frequency values, from minus to plus infinity. This can be thought of in the following way: We think of a non-periodic function as a periodic function with period tending to infinity. This is a valid statement, because as a periodic with period T repeats itself after an interval T , we can say that a non-periodic function repeats itself after $T \rightarrow \infty$. Now, as the time period increases, the spacing between two adjacent frequencies decreases. And

$$f = \lim_{T \rightarrow \infty} \frac{2\pi}{T} = 0.$$

Hence, for a non-periodic function, the spacing between the adjacent frequencies tends to zero, which amounts to the Fourier spectrum going to continuous from discrete. This is exactly what has happened. The Fourier transform $H(\Omega)$ is a continuous function, and in general has a non-zero value for all $\Omega \in \mathbb{R}$.

5.2. Module 2: Lecture 7 Fourier Transforms properties and Fourier series example

5.2.1. Introduction

In this lecture, we will take a few examples of the calculation of the Fourier transform and then build upon the general properties of a Fourier transform like linearity, time shift, modulation and convolution.

Let us start with calculating the Fourier transform of a rectangular pulse.

5.2.2. Fourier transform of a Rectangular Pulse

Let the pulse be symmetric, with height equal to A , going from $-T/2$ to $T/2$ on the time axis. We call this function $h(t)$ or $x(t)$ any one of them as shown in figure below.

Let us denote the Fourier transform by $H(\Omega)$. Then

$$H(\Omega) = \int_{-\infty}^{\infty} h(t)e^{-j\Omega t} dt \quad (5.1)$$

As $h(t)$ is zero except in the time interval $-T/2$ to $T/2$, we can write the above equation as

$$H(\Omega) = \int_{-T/2}^{T/2} Ae^{-j\Omega t} dt = \frac{Ae^{-j\Omega t}}{-j\Omega} \Big|_{-T/2}^{T/2} = \frac{A(e^{-j\Omega T/2} - e^{j\Omega T/2})}{-j\Omega} = \frac{AT(2j \sin(\Omega T/2))}{j\Omega T} \quad (5.2)$$

5. The Fourier Transform

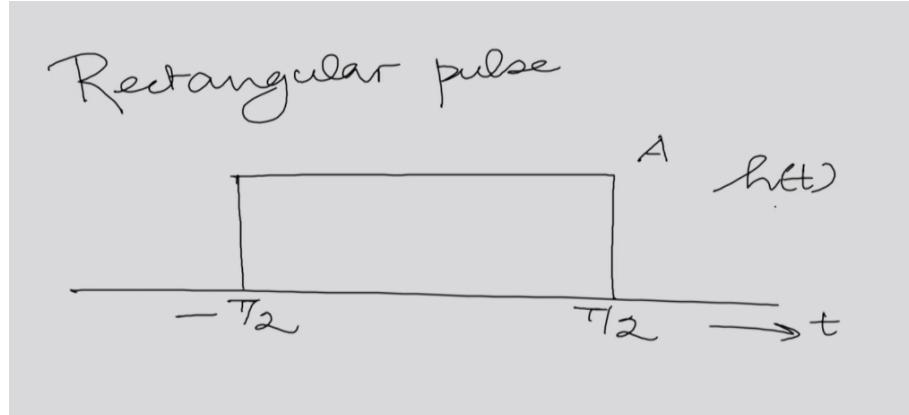


Figure 5.2.: Symmetric Rectangular Pulse

$$H(\Omega) = \frac{AT \sin(\Omega T/2)}{\Omega T/2} \quad (5.3)$$

And we can rewrite it as

$$H(f) = \frac{AT \sin(2\pi f T/2)}{2\pi f T/2} \quad (5.4)$$

Where $f = \frac{\Omega}{2\pi}$ corresponds to the *cycles/second* frequency, and Ω is of course the angular frequency.

Now this expression also can be written as

$$H(\gamma) = \frac{AT \sin(\pi\gamma)}{\pi\gamma} \quad (5.5)$$

Where $\gamma = f * T$

The function $\frac{\sin \pi\gamma}{\pi\gamma}$ is a very special function called a *Sinc Function*. The sinc function in the form presented is often used in the context of signals and systems.

Sketch of the Sinc function

Here are the some properties of the Sinc function.

- It is an even function.
- It has nulls at every integer; so at 1, at 2, at 3 and so on, it will have value equal to zero.
- It tends to 1 as γ tends to 0
- It is a damped sinusoid, i.e. the magnitude of oscillation decreases as γ increases.

Using the above points we can sketch the Sinc function. So the sketch somewhat looks like the figure below.

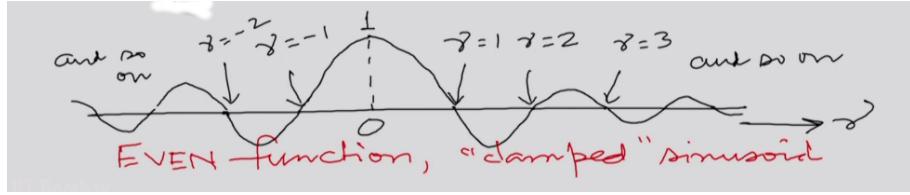


Figure 5.3.: Sinc Function

Important points regarding the Fourier transform of rectangular pulse

The pulse is limited in terms of its occupancy of the time axis. It goes only from $-T/2$ to $+T/2$. On the other hand, the Fourier transform, in principle, last all over the frequency axis. What this means is that if we want to make a function vanish all over the time axis outside a certain finite interval, we have to bring together frequencies of all magnitudes and signs except for a few nulls. All these frequencies need to come together, in appropriate amplitudes, to form this rectangular pulse. Also note that the Fourier transform for this case is purely real.

When the pulse width is halved, the nulls of the Fourier transform are doubly spaced and the central amplitude is halved. This can be checked mathematically. We can see that in the figure below.

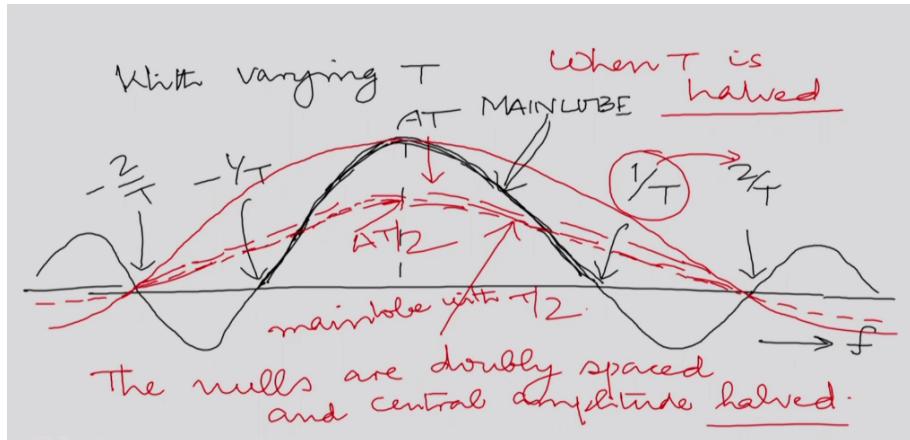


Figure 5.4.: Fourier Transform of Rectangular Pulse at width T and $T/2$

5.2.3. Properties of Fourier Transform

When we change the signal in some way or when we bring signals together, what happens to the corresponding Fourier transform is what we will investigate in this section.

Linearity

Let $h_1(t)$ have the Fourier transform $H_1(\Omega)$ and let $h_2(t)$ have the Fourier transform $H_2(\Omega)$. Then, linearity says

$$\alpha h_1(t) + \beta h_2(t) \xrightarrow{\mathfrak{F}} \alpha H_1(\Omega) + \beta H_2(\Omega)$$

for all $\alpha, \beta \in \mathbb{C}$ and all $h_1(t)$ and $h_2(t)$. Here, \mathfrak{F} means "has the Fourier transform given by".

5. The Fourier Transform

Proof of the linearity

$$\alpha\{H_1(\Omega)\} = \alpha\left\{\int_{-\infty}^{\infty} h_1(t)e^{-j\Omega t}dt\right\} \quad (5.6)$$

$$\beta\{H_2(\Omega)\} = \beta\left\{\int_{-\infty}^{\infty} h_2(t)e^{-j\Omega t}dt\right\} \quad (5.7)$$

Add equation 5.6 and 5.7. We can write it as

$$\{\alpha H_1(\Omega) + \beta H_2(\Omega)\} = \left\{\int_{-\infty}^{\infty} (\alpha h_1(t) + \beta h_2(t))e^{-j\Omega t}dt\right\} \quad (5.8)$$

The Inverse Fourier transform

Inverse Fourier transform can be written as

$$H_1(\Omega) \xrightarrow{\mathfrak{F}^{-1}} \frac{1}{2\pi} \int_{-\infty}^{\infty} H_1(t)e^{j\Omega t}dt \quad (5.9)$$

Note $\Omega = 2\pi f$, $d\Omega = 2\pi df$

Whereupon

$$H_1(f) \xrightarrow{\mathfrak{F}^{-1}} \int_{-\infty}^{\infty} H_1(t)e^{j2\pi ft}dt \quad (5.10)$$

Where Ω = Radians/sec frequency And f = cycles/sec frequency

- We can clearly see that when frequency is in radians/sec then, we will have factor of $1/2\pi$ in the inverse Fourier transform and when frequency is in cycles/sec then the factor of $1/2\pi$ will not be there.
- In a way dealing with cycles/second frequency has some conveniences, we don't need to remember the factor of $1/2\pi$ or we can say, there is a perfect symmetry between the Fourier transform and the Inverse Fourier transform.

Exercise

Prove that **linearity** holds true for the inverse Fourier transform.

5.2.4. Time Shift

what happens to the Fourier transform when we shift the signal in time? If

$$h(t) \xrightarrow{\mathfrak{F}} H(\Omega) \quad (5.11)$$

then

$$h(t - \tau_0) \xrightarrow{\mathfrak{F}} ? \quad (5.12)$$

So the Fourier transform would be

$$H(\Omega) = \int_{-\infty}^{\infty} h(t - \tau_0)e^{-j\Omega t}dt \quad (5.13)$$

Put $(t - \tau_0) = \lambda \Rightarrow t = \tau_0 + \lambda$. So the Fourier transform is

$$H(\Omega) = \int_{-\infty}^{\infty} h(\lambda)e^{-j\Omega(\lambda+\tau_0)}d\lambda = e^{-j\Omega\tau_0} \int_{-\infty}^{\infty} h(\lambda)e^{-j\Omega\lambda}d\lambda \quad (5.14)$$

So,

$$h(t - \tau_0) \xrightarrow{\mathcal{F}} e^{-j\Omega\tau_0} H(\Omega) \text{ or } e^{-j2\pi f\tau_0} H(f) \quad (5.15)$$

Now it can be clearly seen that

- The magnitude of the Fourier transform is unchanged.
- The angle or the phase of the Fourier transform changes.

Modulation

By modulation, we mean multiplying $h(t)$ by a rotating complex number (rotating with an angular velocity of Ω_0). In this case, the Fourier transform is shifted by Ω_0 forward.

$$e^{j\Omega_0 t} h(t) \xrightarrow{\mathcal{F}} ? \quad (5.16)$$

So the Fourier transform would be

$$\int_{-\infty}^{\infty} e^{j\Omega_0 t} h(t) e^{-j\Omega t} dt = \int_{-\infty}^{\infty} h(t) e^{-j(\Omega - \Omega_0)t} dt = H(\Omega - \Omega_0) \quad (5.17)$$

- When we shift a function in time, it causes a modulation in the frequency domain (We remember, it multiplied the Fourier transform by a term $e^{-j\Omega\tau_0}$).
- When we modulate the function in time by multiplying by a rotating complex number, the corresponding Fourier transform shifts in frequency.
- Basically, shift in time becomes modulation in frequency, and modulation in time becomes a shift in frequency.

Convolution Property of Fourier transform

The question to answer here is what will be the Fourier transform of $x_1(t) * x_2(t)$, given that the Fourier transforms of $x_1(t)$ and $x_2(t)$ are $X_1(\Omega)$ and $X_2(\Omega)$ respectively.

$$x_1(t) * x_2(t) = \int_{-\infty}^{\infty} x_1(\tau) x_2(t - \tau) d\tau \quad (5.18)$$

$$x_1(t) * x_2(t) \xrightarrow{\mathcal{F}} ? \quad (5.19)$$

Now let's solve it. The Fourier transform can be written as

$$\int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} x_1(\tau) x_2(t - \tau) d\tau \right\} e^{-j\Omega t} dt \quad (5.20)$$

This is same as

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1(\tau) x_2(t - \tau) e^{-j\Omega t} d\tau dt \quad (5.21)$$

To solve the above double integral we will do a change of variables. Put

$$\alpha = \tau \text{ and } \beta = (t - \tau) \quad (5.22)$$

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \tau \\ t \end{pmatrix} \quad (5.23)$$

5. The Fourier Transform

$$d\tau dt = |\text{Det}(\text{transformation})| d\alpha d\beta \quad (5.24)$$

Hence,

$$d\tau dt = d\alpha d\beta \quad (5.25)$$

Hence, the element of integration also remains the same because the **jacobian** is 1. Now the modified integral is

$$\begin{aligned} & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1(\alpha) x_2(\beta) \exp -j\Omega(\alpha + \beta) d\alpha d\beta \\ &= \int_{-\infty}^{\infty} x_1(\alpha) \left\{ \int_{-\infty}^{\infty} x_2(\beta) e^{-j\Omega(\beta)} d\beta \right\} e^{-j\Omega\alpha} d\alpha \end{aligned} \quad (5.26)$$

This can be written as

$$X_2(\Omega) \int_{-\infty}^{\infty} x_1(\alpha) e^{-j\Omega\alpha} d\alpha$$

So

$$x_1(t) * x_2(t) \xrightarrow{\mathcal{F}} X_1(\Omega) X_2(\Omega) \quad (5.27)$$

Hence a convolution in the natural domain moves to multiplication in the Fourier domain.

5.2.5. Conclusion

In this lecture we discussed about Fourier transform of rectangular pulse which is the “Sinc function” and its properties. We showed some of the properties of Fourier transform and their proofs.

5.3. Module 2: Lecture 8

Convolution Property of the Fourier Transform

5.3.1. Introduction

In this lecture, we will talk about the properties of the Fourier transform and specifically, we see more about the convolution property.

5.3.2. Convolution Theorem

If two signals $x_1(t)$ and $x_2(t)$ both have Fourier transforms, which are $X_1(\Omega)$ and $X_2(\Omega)$ respectively, and if their convolution also has a Fourier transform, then this Fourier Transform is equal to the product $X_1(\Omega) \cdot X_2(\Omega)$.

Mathematically,

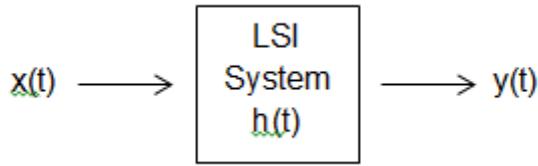
$$x_1(t) \xrightarrow{F} X_1(\Omega) \quad (5.28)$$

$$x_2(t) \xrightarrow{F} X_2(\Omega) \quad (5.29)$$

$$x_1(t) * x_2(t) \xrightarrow{F} X_1(\Omega) \cdot X_2(\Omega) \quad (5.30)$$

So, when we apply it to the context of the Linear Shift Invariance (LSI) system, we get

$$Y(\Omega) = X(\Omega)H(\Omega) \quad (5.31)$$



where,

$$x(t) \xrightarrow{F} X(\Omega) \quad (5.32)$$

$$y(t) \xrightarrow{F} Y(\Omega) \quad (5.33)$$

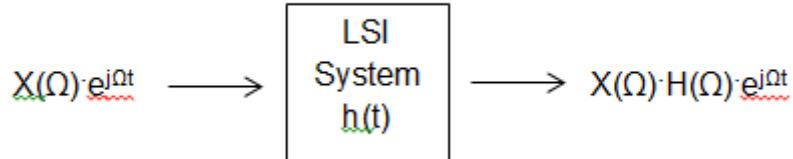
$$h(t) \xrightarrow{F} H(\Omega) \quad (5.34)$$

We said that there is a point-wise decoupling or memory less-ness in the Fourier domain, meaning, the output at angular frequency Ω depends only on the input at angular frequency Ω and the impulse response at angular frequency Ω and none other.

Where does this come from?

$X(\Omega)$ is the component of input at angular frequency Ω . Let us imagine $X(\Omega).e^{j\Omega t}$ as input to the LSI system, where $X(\Omega)$ is a complex number.

Notionally,



$$y(t) = \int_{-\infty}^{\infty} h(\tau).X(\Omega).\exp(j\Omega(t - \tau))d\tau \quad (5.35)$$

$$y(t) = \left(\int_{-\infty}^{\infty} h(\tau).\exp(-j\Omega\tau)d\tau \right) X(\Omega).\exp(j\Omega t) \quad (5.36)$$

$$y(t) = X(\Omega).H(\Omega).\exp(j\Omega t) \quad (5.37)$$

$H(\Omega)$ is the component of the impulse response along $e^{j\Omega t}$. Hence, $X(\Omega).H(\Omega)$ is the component $e^{j\Omega t}$ in $y(t)$ and is equal to $Y(\Omega)$.

$$Y(\Omega) = X(\Omega).H(\Omega) \quad (5.38)$$

This is analogous to an operator acting upon a force in mechanics. If ∇ is the operator and if \vec{F} the force vector such that

$$\vec{F} = F_x \hat{x} + F_y \hat{y} + F_z \hat{z} \quad (5.39)$$

Then

$$\nabla \cdot \vec{F} = \nabla \cdot (F_x \hat{x} + F_y \hat{y} + F_z \hat{z}) \quad (5.40)$$

$$\nabla \cdot \vec{F} = \partial_x F_x \hat{x} + \partial_y F_y \hat{y} + \partial_z F_z \hat{z} \quad (5.41)$$

$$(5.42)$$

5. The Fourier Transform

Similarly we resolve the input into its components along the different Ω 's, i.e. $e^{j\Omega t}$.

$$x(t) = \sum_i X(\Omega_i) \exp(j\Omega_i t) \quad (5.43)$$

$$y(t) = \sum_i X(\Omega_i) \cdot H(\Omega_i) \cdot \exp(j\Omega_i t) \quad (5.44)$$

In a way, the idea is to decouple the action of the LSI system when we think of the input as comprising of different complex exponentials rotating at different angular frequencies, both positive and negative.

5.3.3. What is a transform?

More generally, transform is a change of paradigm. Paradigm means world view. In a system point of view, a signal is a mapping from the independent variable to complex numbers, while a system is a mapping from signals to signals. Whereas a transform is a mapping of the signals and systems altogether to the transformed domain of signals and systems.

5.3.4. Convolution or Multiplication

Doing multiplication, in general, is easier than convolution. To find the convolution, take Fourier transforms of the two signals, multiply them and then take its inverse Fourier transform.

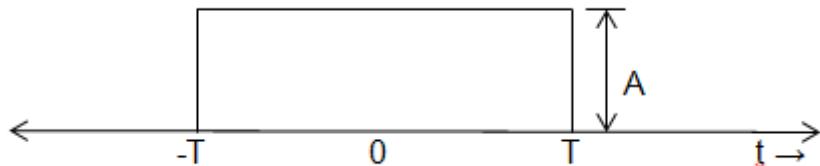
$$x_1(t) \xrightarrow{F} X_1(\Omega) \quad (5.45)$$

$$x_2(t) \xrightarrow{F} X_2(\Omega) \quad (5.46)$$

$$X_1(\Omega) \cdot X_2(\Omega) \xrightarrow{F^{-1}} x_1(t) \cdot x_2(t) \quad (5.47)$$

It is beneficial only if the Fourier transform operations are easier to perform, otherwise it is not always better to go through the Fourier domain.

For example, take $x_1(t) = x_2(t)$ as a rectangular pulse as shown in the figure below. Hence,



$$X_1(\Omega) = X_2(\Omega) = \int_{-\infty}^{\infty} x_1(t) e^{-j\Omega t} dt \quad (5.48)$$

$$= \int_{-T}^{T} A e^{-j\Omega t} dt \quad (5.49)$$

$$= \left(\frac{A}{-j\Omega} e^{-j\Omega t} \right) \Big|_{-T}^{T} \quad (5.50)$$

$$= \frac{2A j \sin(\Omega T)}{j\Omega} \times \frac{T}{T} \quad (5.51)$$

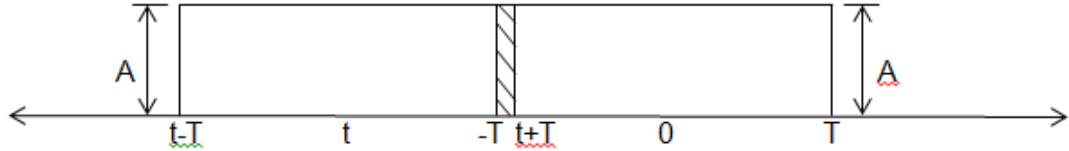
$$= 2AT \frac{\sin \Omega T}{\Omega T} \quad (5.52)$$

Convolving,

$$x_1(t) * x_2(t) = \int_{-\infty}^{\infty} x_1(\tau) \cdot x_2(t - \tau) d\tau \quad (5.53)$$

Now, $x_1(\tau) \cdot x_2(t - \tau)$ is non-zero only when some area of both the pulses coincide and is non-zero for that range only.

This happens only for $t \in [-2T, 2T]$



therefore for $t \in [-2T, 0]$

$$x_1(t) * x_2(t) = \int_{-\infty}^{\infty} x_1(\tau) \cdot x_2(t - \tau) d\tau \quad (5.54)$$

$$= \int_{-T}^{t+2T} A^2 d\tau \quad (5.55)$$

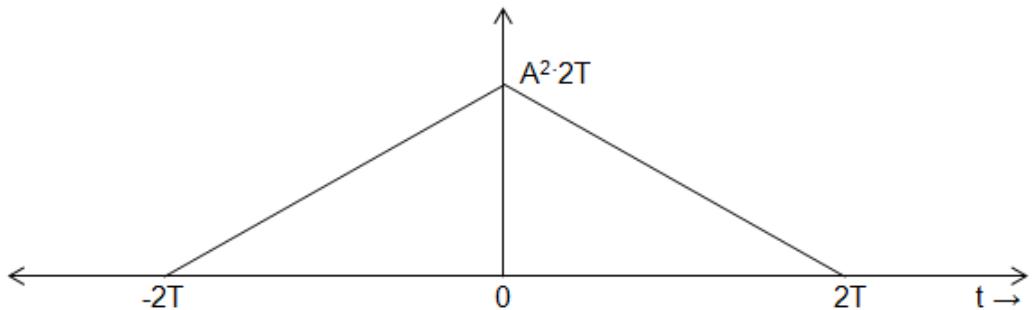
$$= A^2(t + 2T) \quad (5.56)$$

and for $t \in [0, 2T]$

$$x_1(t) * x_2(t) = \int_{-\infty}^{\infty} x_1(\tau) \cdot x_2(t - \tau) d\tau \quad (5.57)$$

$$= \int_{t-T}^T A^2 d\tau \quad (5.58)$$

$$= A^2(2T - t) \quad (5.59)$$



5. The Fourier Transform

Now the Fourier transform of $x_1(t) * x_2(t)$ is

$$= X_1(\Omega)X_2(\Omega)$$

which is equal to

$$\left(2AT\left(\frac{\sin(\Omega T)}{\Omega T}\right)\right)^2$$

5.3.5. The Principle of Duality

We know that

$$X(\Omega) = \int_{-\infty}^{\infty} x(t)e^{-j\Omega t} dt$$

and

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\Omega)e^{j\Omega t} d\Omega$$

Now, in the second integral, let's interchange 't' and ' Ω '. Hence,

$$x(\Omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(t)e^{j\Omega t} dt$$

Hence,

$$x(-\Omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(t)e^{-j\Omega t} dt$$

Hence,

$$2\pi[x(-\Omega)] = \int_{-\infty}^{\infty} X(t)e^{-j\Omega t} dt$$

This shows that, if

$$x(t) \xrightarrow{\mathcal{F}} X(\Omega)$$

then

$$X(t) \xrightarrow{\mathcal{F}} 2\pi[x(-\Omega)] \quad (5.60)$$

This is the principle of duality for the Fourier transform.

Let us take the example of the same rectangular pulse which we considered in the earlier section. The Fourier transform of the rectangular pulse is given by

$$X(\Omega) = 2AT \frac{\sin(\Omega T)}{\Omega T}$$

Let us invoke duality here. We will replace T by W for convenience of notation. Hence,

$$X(t) = 2AW \frac{\sin(Wt)}{Wt} \xrightarrow{\mathcal{F}} 2\pi[x(-\Omega)]$$

Here, $2\pi x(\Omega)$ is the rectangular pulse with width $2W$ and height $2\pi A$. As the pulse is symmetric, $x(-\Omega) = x(\Omega)$.

We can see why this is a very useful property. Say you want to evaluate the convolution of $\sin(Wt)/Wt$. It is very difficult to evaluate it using the direct expression of the convolution. But, we can invoke the principle of duality here. We know that the Fourier transform of a convolution is the product of their Fourier transforms. Hence, applying this to $X(t)$,

$$X(t) * X(t) \xrightarrow{\mathcal{F}} [2\pi x(\Omega)]^2$$

Now, it is very easy to evaluate the right hand side. It is in fact the same symmetric rectangular pulse, but with height $(2\pi A)^2$. Hence, it is equal to $(2\pi)^2 A x(\Omega)$. Therefore, we have,

$$X(t) * X(t) \xrightarrow{\mathcal{F}} (2\pi)^2 A x(\Omega) \quad (5.61)$$

Now, we know from the principle of duality (eq. (33)), that

$$X(t) \xrightarrow{\mathcal{F}} 2\pi[x(-\Omega)]$$

Comparing this with eq. (34), we get,

$$X(t) * X(t) = 2\pi A X(t) = 2\pi A \left(2AW \frac{\sin(Wt)}{Wt} \right)$$

Hence, the evaluation of the convolution of $\sin(Wt)/Wt$ and other complicated functions with themselves can become a lot easier and straightforward if we invoke the principle of duality.

5.4. Module 2: Lecture 9

Multiplication Theorem and Parseval's Theorem

5.4.1. Introduction

We have seen in the last few sessions, the property of duality and the convolution – multiplication parallel. We will now see the implications of duality, deriving what we call ‘Multiplication theorem’ and then its application called the ‘Parseval’s theorem’.

5.4.2. Multiplication

Let us take $x_1(t)$ and $x_2(t)$ with Fourier transforms $X_1(\Omega)$ and $X_2(\Omega)$ respectively. We derived earlier the Fourier transform of $x_1(t) * x_2(t)$ to be $X_1(\Omega)X_2(\Omega)$. We shall now try to find a general expression for the Fourier transform of the product $x_1(t)x_2(t)$.

We have,

$$x_1(t) \xrightarrow{\mathcal{F}} X_1(\Omega) \quad (5.62)$$

$$x_2(t) \xrightarrow{\mathcal{F}} X_2(\Omega) \quad (5.63)$$

$$x_1(t) * x_2(t) \xrightarrow{\mathcal{F}} X_1(\Omega)X_2(\Omega) \quad (5.64)$$

Applying duality on (3) gives

$$X_1(t)X_2(t) \xrightarrow{\mathcal{F}} 2\pi(x_1 * x_2)(-\Omega)$$

Also by duality,

$$X_1(t) \xrightarrow{\mathcal{F}} 2\pi x_1(-\Omega)$$

$$X_2(t) \xrightarrow{\mathcal{F}} 2\pi x_2(-\Omega)$$

$2\pi x_1(-\Omega) * x_2(-\Omega)$, which is the Fourier transform of $X_1(t)X_2(t)$, can be re-written as,

$$\frac{1}{2\pi}(2\pi x_1(-\Omega) * 2\pi x_2(-\Omega))$$

Notice that $2\pi x_1(-\Omega)$ is the Fourier transform of $X_1(t)$ and $2\pi x_2(-\Omega)$ is the Fourier transform of $X_2(t)$. Therefore we can see that the Fourier transform of the product $X_1(t)X_2(t)$ is the convolution of Fourier transforms of the individual time domain functions multiplied by a factor of $\frac{1}{2\pi}$.

5. The Fourier Transform

Theorem

If

$$\begin{aligned} y_1(t) &\xrightarrow{\mathcal{F}} Y_1(\Omega) \\ y_2(t) &\xrightarrow{\mathcal{F}} Y_2(\Omega) \end{aligned}$$

Then the Fourier transform the product (provided all the Fourier transforms exist) is,

$$y_1(t)y_2(t) \xrightarrow{\mathcal{F}} \frac{1}{2\pi}(Y_1(\Omega) * Y_2(\Omega)) \quad (5.65)$$

Multiplication of one signal by another can be thought of as using one signal to scale or modulate the amplitude of the other, and consequently, the multiplication of two signals is often referred to as 'amplitude modulation'. For this reason, (4) is sometimes referred to as the modulation property.

Special Case of Multiplication Theorem

We have seen from the multiplication theorem, if $y_1(t)$ and $y_2(t)$ have the following Fourier transforms,

$$\begin{aligned} y_1(t) &\xrightarrow{\mathcal{F}} Y_1(\Omega) \\ y_2(t) &\xrightarrow{\mathcal{F}} Y_2(\Omega) \end{aligned}$$

then

$$y_1(t)y_2(t) \xrightarrow{\mathcal{F}} \frac{1}{2\pi}Y_1(\Omega) * Y_2(\Omega)$$

Let us find the Fourier transform of $y_1(t)\overline{y_2(t)}$: We will first find out the Fourier transform of $\overline{y_2(t)}$ and then get the Fourier transform of $y_1(t)$ in two different ways:

1. Using the multiplication theorem.
2. By the general definition of Fourier transform of any given function.

First, the Fourier transform of $\overline{y_2(t)}$: The inverse Fourier transform of a given function $X(\Omega)$ is given by

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\Omega) e^{j\Omega t} d\Omega$$

Taking the complex conjugate of the above,

$$\overline{x(t)} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \overline{X(\Omega)} e^{-j\Omega t} d\Omega$$

Transforming Ω with $-\alpha$,

$$\overline{x(t)} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \overline{X(-\alpha)} e^{j\alpha t} d\alpha \quad (5.66)$$

From equation (5) we can observe that the Fourier transform of $\overline{x(t)}$ is $\overline{X(-\alpha)}$. Therefore

$$\overline{y_2(t)} \xrightarrow{\mathcal{F}} \overline{Y_2(-\Omega)}$$

From multiplication theorem, we can get the Fourier transform of the product $y_1(t)\overline{y_2(t)}$

$$y_1(t)\overline{y_2(t)} \xrightarrow{\mathcal{F}} \frac{1}{2\pi}Y_1(\Omega) * \overline{Y_2(-\Omega)}$$

We can also obtain the Fourier transform using the general definition, i.e

$$y_1(t)\overline{y_2(t)} \xrightarrow{\mathcal{F}} \int_{-\infty}^{\infty} y_1(t)\overline{y_2(t)}e^{-j\Omega t}dt$$

The Fourier transforms of $y_1(t)\overline{y_2(t)}$ obtained by either method should be identical, and hence we can write,

$$\begin{aligned} \frac{1}{2\pi} Y_1(\Omega) * \overline{Y_2(-\Omega)} &= \int_{-\infty}^{\infty} y_1(t)\overline{y_2(t)}e^{-j\Omega t}dt \\ \frac{1}{2\pi} \int_{-\infty}^{\infty} Y_1(\Omega - \lambda)\overline{Y_2(-\lambda)} d\lambda &= \int_{-\infty}^{\infty} y_1(t)\overline{y_2(t)}e^{-j\Omega t}dt \end{aligned}$$

In the above equation, put $\Omega = 0$, the identity becomes

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} Y_1(-\lambda)\overline{Y_2(-\lambda)}d\lambda = \int_{-\infty}^{\infty} y_1(t)\overline{y_2(t)}dt$$

In the left hand side integral above, transform using $-\lambda \rightarrow \beta$, then $d\lambda \rightarrow -d\beta$, and as λ goes from $-\infty$ to $+\infty$, β goes from $+\infty$ to $-\infty$. The equation becomes

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} Y_1(\beta)\overline{Y_2(\beta)}d\beta = \int_{-\infty}^{\infty} y_1(t)\overline{y_2(t)}dt$$

Observe that the right hand side is essentially the inner product of $y_1(t)$ and $y_2(t)$ and the left hand side is the inner product of $Y_1(\Omega)$ and $Y_2(\Omega)$ multiplied by a factor of $\frac{1}{2\pi}$. Essentially, arrived at an equivalence between inner products in time domain and inner product in frequency domain. This is called the Parseval's theorem.

Theorem

If $y_1(t)$ and $y_2(t)$ have respectively their Fourier transforms $Y_1(\Omega)$ and $Y_2(\Omega)$, then the inner product of $y_1(t)$ and $y_2(t)$ is $\frac{1}{2\pi}$ times the inner product of $Y_1(\Omega)$ and $Y_2(\Omega)$

$$\int_{-\infty}^{\infty} y_1(t)\overline{y_2(t)}dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} Y_1(\Omega)\overline{Y_2(\Omega)}d\Omega$$

5.5. Module 2: Lecture 10

Spectral Density

5.5.1. Introduction

In the previous lecture, we derived the multiplication property of the Fourier Transform which subsequently led to the derivation of Parseval's Theorem. In this lecture, we will see more on the interpretation of the Parseval's theorem as the invariance of the inner product of two signals calculated in time and angular frequency (or Fourier) domain. We will also define the energy of a signal and relate it with the Parseval's Theorem and introduce the notion of Spectral Density of a signal. Having done that, we will introduce the differentiation property of the Fourier Transform and its dual which will help us in calculating Fourier transforms of many more signals from the already known simpler Fourier transforms.

5. The Fourier Transform

5.5.2. Invariance of Inner Product on Change of Basis

The property

$$\int_{-\infty}^{\infty} x(t)\overline{y(t)}dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\Omega)\overline{Y(\Omega)}d\Omega$$

has a deeper implication when we think of this in terms of the inner product of x and y . Recall from vector algebra that in two dimensions, we can write any vector \vec{v} as a linear combination of any two orthonormal basis vectors. Consider two such vectors, and two different sets of orthonormal vectors, such that

$$\vec{v}_1 = v_{11}\hat{u}_1 + v_{12}\hat{u}_2 = v_{13}\hat{u}_3 + v_{14}\hat{u}_4$$

and

$$\vec{v}_2 = v_{21}\hat{u}_1 + v_{22}\hat{u}_2 = v_{23}\hat{u}_3 + v_{24}\hat{u}_4$$

Now, the inner product or the dot product between the two vectors is the magnitude of one vector times the projection of the second vector along the first vector. The important thing to note here is that the inner product is a property of the vectors in themselves. This inner product, by definition, *does not depend on which basis we choose to express them*. Thus,

$$\vec{v}_1 \cdot \vec{v}_2 = v_{11}v_{21} + v_{12}v_{22} = v_{13}v_{23} + v_{14}v_{24}$$

This is exactly what the meaning of Parseval's theorem is. In the time domain, the basis vectors used, as you can remember from the first model, are the unit impulses:

$$x(t) = \int_{-\infty}^{+\infty} x(\lambda)\delta(t - \lambda) d\lambda$$

Whereas in the frequency domain, the basis vectors are the rotating complex numbers or phasors:

$$X(\Omega) = \int_{-\infty}^{+\infty} x(\lambda)e^{-j\Omega\lambda} d\lambda$$

But the inner product between two signals remains independent of the basis, which is what the Parseval's theorem states.

$$\int_{-\infty}^{\infty} x(t)\overline{y(t)}dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\Omega)\overline{Y(\Omega)}d\Omega$$

The factor of 2π arises due to the same reason why it arrived in the inverse Fourier transform, which is the normalization of the complex exponential.

5.5.3. Energy of a Signal

The energy E_x of a continuous time signal $x(t)$ is defined as

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

Relating Energy of a Signal and Parseval's Theorem

Consider two continuous time signals $x(t)$ and $y(t)$ with Fourier Transforms $X(\Omega)$ and $Y(\Omega)$ respectively. Then Parseval's theorem states that,

$$\int_{-\infty}^{\infty} x(t)\overline{y(t)}dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\Omega)\overline{Y(\Omega)}d\Omega$$

where $(\cdot)^*$ denotes the complex conjugate of the corresponding signal.

Now, if we substitute $y(t) = x(t)$ in the above relation, we get,

$$\int_{-\infty}^{\infty} x(t)\overline{x(t)}dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\Omega)\overline{X(\Omega)}d\Omega$$

Since, $x(t)\overline{x(t)} = |x(t)|^2$ and $X(\Omega)\overline{X(\Omega)} = |X(\Omega)|^2$ we get,

$$\int_{-\infty}^{\infty} |x(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |X(\Omega)|^2 d\Omega$$

From the above relation, we see that the energy of a signal can also be calculate using the formula,

$$E_x = \frac{1}{2\pi} \int_{-\infty}^{\infty} |X(\Omega)|^2 d\Omega$$

Energy Spectral Density

We have seen that if a signal $x(t)$ has a Fourier Transform $X(\Omega)$, its energy can be calculated from the angular frequency domain using the formula,

$$E_x = \frac{1}{2\pi} \int_{-\infty}^{\infty} |X(\Omega)|^2 d\Omega$$

Integrating $|X(\Omega)|^2$ over the entire angular frequency axis and multiplying the result by $\frac{1}{2\pi}$ gives us the energy of the signal.

Thus $|X(\Omega)|^2$ conveys the distribution of energy along the angular frequency axis. Hence, the integrand $|X(\Omega)|^2$ in the above formula is of significance and is called the Energy Spectral Density of the signal $x(t)$.

Energy Spectral Density and Linear Shift-Invariant Systems

Consider a linear shift invariant system having an impulse response $h(t)$ and let $y(t)$ be the output of this system when an input $x(t)$ is applied to it.

Then, we have

$$y(t) = x(t) * h(t)$$

where $*$ denotes the convolution operator.

Assume that the Fourier transforms of $h(t)$ and $x(t)$. Let $H(\Omega)$ and $X(\Omega)$ be their respective Fourier Transforms.

5. The Fourier Transform

Then the Fourier Transform $Y(\Omega)$ of the signal $y(t)$ is given by.

$$Y(\Omega) = X(\Omega)H(\Omega)$$

Therefore,

$$|Y(\Omega)|^2 = |X(\Omega)|^2|H(\Omega)|^2$$

Note that $|Y(\Omega)|^2$, $|X(\Omega)|^2$ and $|H(\Omega)|^2$ are essentially the energy spectral densities of the signals $y(t)$, $x(t)$ and the impulse response $h(t)$ respectively.

Thus, the energy spectral density of the output is equal to the energy density of the input multiplied by the energy density of the frequency response.

5.5.4. The Differentiation Property of the Fourier Transform

Consider a continuous time signal $x(t)$ which has a Fourier transform $X(\Omega)$ then we can write,

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\Omega) e^{j\Omega t} d\Omega$$

Differentiating on both sides with respect to ' t ' we get,

$$\frac{dx(t)}{dt} = \frac{d}{dt} \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} X(\Omega) e^{j\Omega t} d\Omega \right)$$

Taking the derivative under the integral we get,

$$\frac{dx(t)}{dt} = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\Omega) \frac{de^{j\Omega t}}{dt} d\Omega$$

Therefore,

$$\frac{dx(t)}{dt} = \frac{1}{2\pi} \int_{-\infty}^{\infty} (X(\Omega) j\Omega) e^{j\Omega t} d\Omega$$

Comparing the above equation with the definition of the inverse Fourier Transform, we see that for the signal $x(t)$ has the Fourier Transform $j\Omega X(\Omega)$.

Thus, the differentiation property of the Fourier transform states that if a continuous variable signal $x(t)$ has the Fourier Transform $X(\Omega)$, then the Fourier transform of $\frac{dx(t)}{dt}$ is $j\Omega X(\Omega)$.

Interpretation of the differentiation property

The differentiation property of the Fourier Transform can be interpreted in the following way: If you differentiate a signal $x(t)$ with respect to t , then the Fourier transform $X(\Omega)$ gets phase shifted by 90 degrees and scaled by Ω .

Duality and the differentiation property

We can see that there are two operators involved in the differentiation property - differentiation in time domain and pointwise multiplication by $j\Omega$ in angular frequency domain. But, what changes happens to the signal $x(t)$ when we differentiate the Fourier Transform $X(\Omega)$ with respect to Ω in angular frequency domain? Let's find out.

We know that,

$$X(\Omega) = \int_{-\infty}^{\infty} x(t)e^{-j\Omega t} dt$$

Differentiating both the sides with respect to Ω we get,

$$\frac{dX(\Omega)}{d\Omega} = \frac{d}{d\Omega} \int_{-\infty}^{\infty} x(t)e^{-j\Omega t} dt$$

Taking the derivative under the integral we get,

$$\frac{dX(\Omega)}{d\Omega} = \int_{-\infty}^{\infty} x(t) \frac{de^{-j\Omega t}}{d\Omega} dt$$

Therefore,

$$\frac{dX(\Omega)}{d\Omega} = \int_{-\infty}^{\infty} -jtx(t)e^{-j\Omega t} dt$$

Comparing the above equation with the definition of the Fourier Transform we see that Fourier transform of $-jtx(t)$ is $\frac{dX(\Omega)}{d\Omega}$.

Thus, differentiation of Fourier Transform $X(\Omega)$ with respect to Ω in angular frequency domain leads to pointwise multiplication of signal $x(t)$ by $-jt$ in the time domain. Hence, the above property is the dual of the differentiation property.

Differentiation in one domain leads to pointwise multiplication in the other domain and pointwise multiplication in one domain leads to differentiation in the other domain. This is the duality property of the Fourier Transform.

Part III.

Sampling

6. chapter

6.1. Lecture 1: Introduction: The Basic Theme of Module Three

6.1.1. Introduction

6.1.2. Recap of EE210.1x

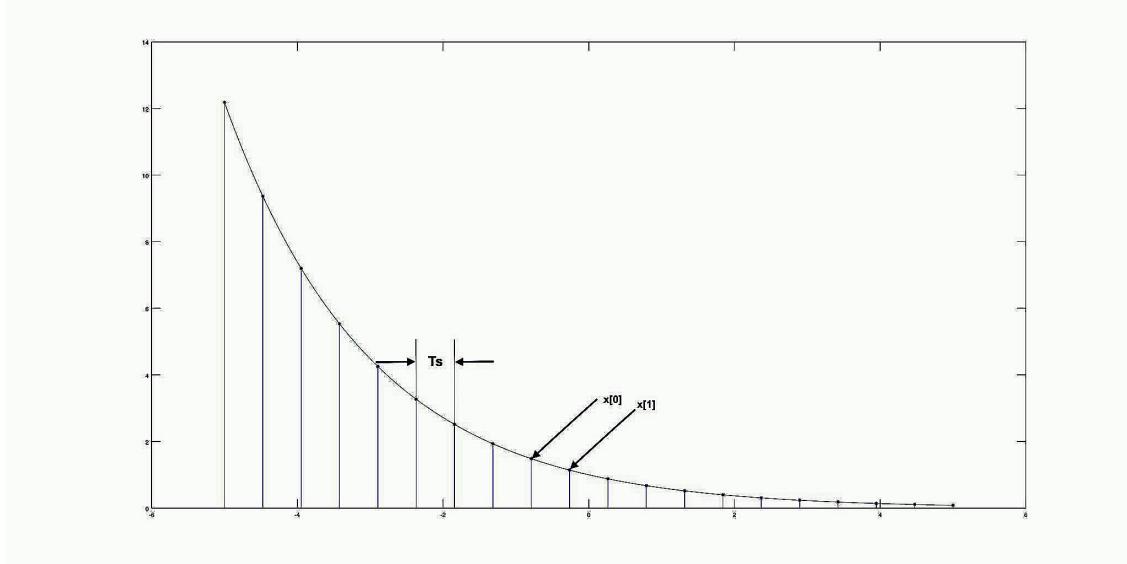
This is the second part of the two part course series on Signals and Systems. In the previous part we studied signals and systems in the time and the frequency domain. The time domain is the natural domain of signals and systems in the sense that they are observable directly in this domain in practical life. The frequency domain however is as important to the study of signals and systems as the time domain because it provides us a way to represent signals in terms of sinusoids (or equivalently complex exponentials) which are more conducive to analysis as compared to the original signal. Specifically, a time limited signal could be represented by its Fourier series expansion after performing a periodic extension of the signal on the entire axis. Thus the signal can be represented using a (possibly infinite) set of discrete frequencies. In the case of a signal extending over the entire time axis, this is not possible and we need to consider the entire continuum of frequencies in the general case.

In the first course, however, we always considered continuous and discrete signals and systems independently though in parallel. Though this is a good way to introduce the two apparently disjoint and unrelated concepts, we must note that we need to, at some point, study both of them together as there exist very fundamental and very practical connections between the two. We start off by noting again that a time limited signal can be represented by equi spaced discrete frequencies in the frequency domain. Thus a continuous signal in one domain (the time domain) is equivalent to a discrete signal in the other domain (the frequency domain). This represents the relation of the two types of signals and systems we studied previously.

6.1.3. Sampling

The relation between these can also be arrived at or made apparent from a much more practical phenomenon. Have you ever seen a LP record? It is a big black disk that is played using a gramophone. We no longer see people with these. As is thus obvious, the LP record has been phased out to be replaced by storage media like the optical CD. Have you ever thought why this so happened? The answer lies in the format of storage of audio on these different storage media. In the record, sound is stored as a function of time as a continuous variable, like sound is actually found in the real world. However, in an optical disk, sound is recorded as a function of time as a discrete variable. We compute the amplitude of the sound at equal intervals of time and store these obtained values. We *sample* the sound at fixed intervals of time (equivalently at a fixed *sampling frequency*) and store the obtained discrete signal. Such discretising of the media has advantages in

6. chapter



terms of robustness against noise, ease of processing and ease storage. But that is another story. Well, purists disagree of course :P!

In fig. 1, we see a continuous time signal $x(t)$ being sampled at a frequency of $1/T_s$ where T_s is the length of the *sampling interval*. The origin (the point $t = 0$) of the resulting discrete signal can be assigned as per convenience.

We obtain, then, from the continuous signal, a discrete signal which, in some ways, is similar to the original continuous signal; it is a replica of the continuous signal. Music lovers might be like, "Wait a minute! Aren't we losing out on a lot of information in doing that? Won't my music sound bad?" We, however, resort to some clever techniques so that this loss is not perceptible to the human senses. And we can always resort to a higher sampling frequency. Thus, in the case of an audio CD, the sampling rate is 44,100 Hz i.e. we sample the audio signal 44,100 times in a second! We hope, thus, that we do not lose out a lot on the human experience of the music (or the image or video. Yes even the images we deal with these days are discretised. In what way do you think is this done?).

In this module, we set out to answer this basic about sampling a continuous time signal. We recognise that it is necessary and convenient to sample the signal as mentioned above. However, we must ask ourselves if the sampling we perform is even meaningful; meaningful in the sense that this sampling does not cost us a lot. We might lose information when sampling. Is this information loss tolerable or can we process the signals (both the sampled and the original) so that the result is acceptable. Or should we just increase the sampling frequency? While pondering over these, we must remember the enormous benefits sampling provides us. Thus it is a trade off and we would attempt to provide an analytical approach to solve this dilemma in the future lectures. While studying this, we explore the aforementioned connection between discrete and continuous signals and systems.

6.1.4. A Priori Information

When going from a continuous signal to a sampled version of it, we noted above that we might end up loosing some information contained in the original signal i.e. the signal reconstructed from the sampled signal may not resemble the original signal as much as we might want it to. The central idea in determining if this is the case, if reconstructing the original signal is simple or hard is the idea of apriori information. Apriori information is the information we have about the signal before we start the process of reconstruction. We illustrate this idea with a series of examples below.

Some Examples

Some Easy Cases

Consider a constant continuous signal $x(t) = \alpha, \alpha \in \mathbb{R}$. We ask the question in the context of this signal now. Can we sample and reconstruct this signal? If so, how do we go about doing this? If we have the a priori information that the signal is a constant signal, then we just need to know the value of α to have complete knowledge about the signal so that we can reconstruct it fully. This can be achieved by sampling the signal at just one value of t , say t_0 . As $x(t_0) = \alpha$ we get the value of α as required.

Consider now the situation where we know beforehand that the signal under consideration is zero everywhere. Then, in order to reconstruct it, we need no samples at all. Even in the case of a non trivial signal like a sinusoid given by $x(t) = 5\cos(4t + 2)$, if know that the signal is $x(t) = 5\cos(4t + 2)$, we do not need any samples to reconstruct the signal.

These examples bring out an important idea. The more a priori information we have, the less is the amount of effort needed to sample and reconstruct the signal.

Reconstructing an Exponential Signal

Consider the case where the a priori information we have is that the signal is an exponential signal. Thus the signal is of the form $x(t) = A_0 e^{\alpha t}$. In order to sample and reconstruct this signal, we observe that we need two samples, say at times t_1 and t_2 . The values of the samples will thus be $x(t_1)$ and $x(t_2)$ respectively. Then, we obtain the following two equations.

$$x(t_1) = A_0 e^{\alpha t_1} \quad (6.1)$$

$$x(t_2) = A_0 e^{\alpha t_2} \quad (6.2)$$

Dividing the (1) by (2) and taking the natural logarithm on both sides, we obtain

$$\alpha = \frac{\ln(x(t_2)) - \ln(x(t_1))}{t_2 - t_1}$$

Substituting this value of α in (1), we obtain,

$$A_0 = x(t_1) e^{-\alpha t_1}$$

Thus, samples at two finite times enable us to reconstruct the signal entirely with the a priori information we have.

6. chapter

Making the Problem Harder: A Sum of Exponentials

Suppose we just know that the signal is finite sum of exponentials. Thus, if $M \in \mathbb{Z}$,

$$x(t) = A_1 e^{\alpha_1 t} + A_2 e^{\alpha_2 t} + \dots + A_M e^{\alpha_M t}$$

If M is unknown too, this problem is very hard to solve. Thus, for the purposes of illustration, we assume that M is known. In that case, we have $2M$ unknowns, $\alpha_1, \alpha_2, \dots, \alpha_M$ and A_1, A_2, \dots, A_M . Thus, we need $2M$ samples to be able to reconstruct the signal entirely. Let the samples be taken at times t_1, t_2, \dots, t_{2M} . In this case, we get the following system of equations,

$$A_1 e^{\alpha_1 t_1} + A_2 e^{\alpha_2 t_1} + \dots + A_M e^{\alpha_M t_1} = x(t_1)$$

$$A_1 e^{\alpha_1 t_2} + A_2 e^{\alpha_2 t_2} + \dots + A_M e^{\alpha_M t_2} = x(t_2)$$

$$\vdots$$

$$A_1 e^{\alpha_1 t_{2M}} + A_2 e^{\alpha_2 t_{2M}} + \dots + A_M e^{\alpha_M t_{2M}} = x(t_{2M})$$

As can be seen however, this system of equations is highly non-linear in the unknowns and hence it is not exactly clear how a solution might be obtained. At this stage, we simplify the problem further. We assume that $\alpha_1, \alpha_2, \dots, \alpha_M$ are known. In this case, M samples suffice and we get the following system of equations.

$$A_1 e^{\alpha_1 t_1} + A_2 e^{\alpha_2 t_1} + \dots + A_M e^{\alpha_M t_1} = x(t_1)$$

$$A_1 e^{\alpha_1 t_2} + A_2 e^{\alpha_2 t_2} + \dots + A_M e^{\alpha_M t_2} = x(t_2)$$

$$\vdots$$

$$A_1 e^{\alpha_1 t_M} + A_2 e^{\alpha_2 t_M} + \dots + A_M e^{\alpha_M t_M} = x(t_M)$$

Then, this is a system of linear equations in A_1, A_2, \dots, A_M . The fact that this systems is obtained by sampling a sum of exponentials guarantees that the M equations are all independent and that a solution exists. This solution is also easily computable.

One can also assume, however, that the A_1, A_2, \dots, A_M are known and that $\alpha_1, \alpha_2, \dots, \alpha_M$ are unknowns. We get the same system of equations as in the previous case. How can one solve this system to obtain $\alpha_1, \alpha_2, \dots, \alpha_M$? We leave this as an exercise to the students.

6.2. Lecture 2: Non-uniqueness due to sampling

6.2.1. Introduction

In previous lecture, we used a-priori information for filling information between different samples. For example take a exponential signal

$$x(t) = A e^{\alpha t}$$

We can do a unique reconstruction by just doing two measurements and filling everything in between by using a priori information of $x(t)$. Here we will look at what happens when we don't have a-priori information for reconstruction.

6.2.2. Non-uniqueness due to lack of a-priori Information

Let's sketch the reconstruction of $x(t)$ for two measurements

$$x(t_1) = Ae^{\alpha t_1}$$

$$x(t_2) = Ae^{\alpha t_2}$$

and $t_1 \neq t_2$. If we have a-priori form we can fill everything in between t_1 and t_2 and elsewhere using that information (black line in Figure 6.1). Now suppose we know the same samples $x(t_1)$ and $x(t_2)$ and a-priori form is not given, then how we will proceed with reconstruction. The problem is not that there are no possibilities but that there are too many of them. Some of them are shown in Figure 6.1 along with the reconstruction when the given form is exponential.

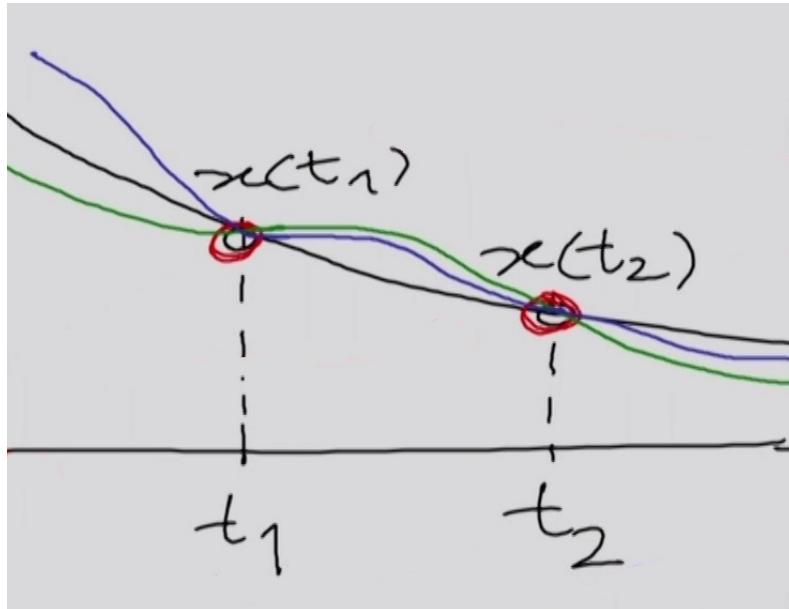


Figure 6.1.: Non-unique reconstruction due to lack of a-priori Information

We can say the problem of sampling and reconstruction is similar to solving a crossword. We can think of clues as a-priori information and letters provided as samples. Without any a-priori information we can fill crossword any way we want but with clues we are restricted in our choice of filling up the crossword. Crosswords are generally designed such that there is a unique solution. Similarly in case of sampling and reconstruction a strong a-priori information restricts us to a unique reconstruction.

Degree of non-uniqueness in reconstruction

When we know nothing about the form of signal then we see there is an infinity of possibilities. To answer how big is this infinity we must first see different classes of infinity. There is an infinity of natural numbers, infinity of integers, positive even/odd numbers etc. which are all same. Then there is an infinity of Real numbers which is

6. chapter

bigger or higher than the infinity of integers in the sense that we can't put real numbers in one to one correspondence to integers without missing some real numbers. In case of reconstruction we need to define a value at every point. Some finite number of points have fixed value which are basically the samples but there is an infinity of points which don't have fixed value and can take any value in the absence of a-priori information. This infinity of possible reconstruction is thus even higher than infinity of real numbers. Now we can understand the degree of non-uniqueness of this problem.

How uniquely we can solve this problem of reconstruction depends on whether the given a-priori information completely complements the samples we have. We have previously seen that using Fourier transform we can represent a signal in terms of sinusoid. Let us take a sine wave and sample it and see what ambiguities we create.

6.2.3. Sampling Sinusoid

Let us take a single sine wave and sample it uniformly i.e sampling with equal spacing between adjacent points. Now we need to find what are all sine waves which have these samples. The problem of ambiguity arises because we can miss an integer number of cycles between two samples. We can also arrive at a wrong edge. So with each possibility of losing a integer number of cycle one can also arrive at right edge or wrong edge. Let's analyse this mathematically by taking

$$A_o \cos(\Omega_o t + \phi_o)$$

as the form of sinusoid where Ω_o is the angular frequency and T_o is the time period

$$\Omega_o = \frac{2\pi}{T_o}$$

Now the process of sampling creates many 'ghost' or 'monster' frequencies which have the same samples at the same points. Let's denote these frequencies by Ω_{kj} where $j(j = 1, 2)$ corresponds to two possibilities of correct or wrong edge.

$$x(t) = A_o \cos(\Omega_{kj} t + \phi_{kj})$$

$$k = 1, 2, 3, \dots$$

When $x(t)$ is sampled at time nT_s where n belongs to set of integers, we get

$$x[nT_s] = A_o \cos(\Omega_o n T_s + \phi_o)$$

The non-uniqueness comes from being able to add a phase to this. Let's add $\pm 2\pi kn$ to the phase where n is the sampling instant and k belongs to set of positive integer.

$$\begin{aligned} & A_o \cos(\Omega_o n T_s \pm 2\pi kn + \phi_o) \\ &= A_o \cos(n T_s (\Omega_o \pm \frac{2\pi}{T_o} k) + \phi_o) \\ &= A_o \cos(2\pi n T_s (\frac{1}{T_o} \pm \frac{k}{T_s}) + \phi_o) \end{aligned}$$

We need to have more than one sample in a cycle. We can see importance of this by taking a extreme example where we take just one sample per cycle and at the same location in every cycle. We will get same sample in every cycle. We cant even tell whether it is a DC

6.2. Lecture 2: Non-uniqueness due to sampling

or a sinusoid. Thus we can say that time interval between the samples must be less than cycle time.

$(\frac{1}{T_o} \pm \frac{k}{T_s})$ represents all possible Hertz frequencies. Let's take the possibility of $k = 1$

$$(\frac{1}{T_o} - \frac{1}{T_s}) < 0$$

as $T_s < T_o$ but we don't want negative frequencies. Let's go back to original expression and see how we can correct this.

$$\begin{aligned} A_o \cos(2\pi n T_s (\frac{1}{T_o} \pm \frac{k}{T_s}) + \phi_o) \\ = A_o \cos \pm (2\pi n T_s (\frac{1}{T_o} \pm \frac{k}{T_s}) + \phi_o) \end{aligned}$$

If we take $+$ we get the same expression and if we take $-$ we get

$$= A_o \cos(-2\pi n T_s (\mp \frac{k}{T_s} + \frac{1}{T_o}) - \phi_o)$$

we see that the phase is reversed.

Ghost Frequencies: An Example

Let's take the case of $k = 1$ again. Take $+$ for $(\frac{1}{T_s} + \frac{1}{T_o})$

$$A_o \cos(2\pi n T_s (\frac{1}{T_o} + \frac{1}{T_s}) + \phi_o)$$

Take $-$ for $(-\frac{1}{T_s} + \frac{1}{T_o})$

$$A_o \cos(2\pi n T_s (\frac{1}{T_s} - \frac{1}{T_o}) - \phi_o)$$

Take $T_s = \frac{T_o}{4}$

$$\frac{1}{T_o} + \frac{1}{T_s} = \frac{5}{T_o}$$

$$\frac{1}{T_s} - \frac{1}{T_o} = \frac{3}{T_o}$$

Let's analyse this situation graphically. Take $T_o = \frac{\pi}{2}$. In Figure 6.2 Black reconstruction is the true sinusoid and red is the sinusoid with ghost frequency $= \frac{5}{T_o}$.

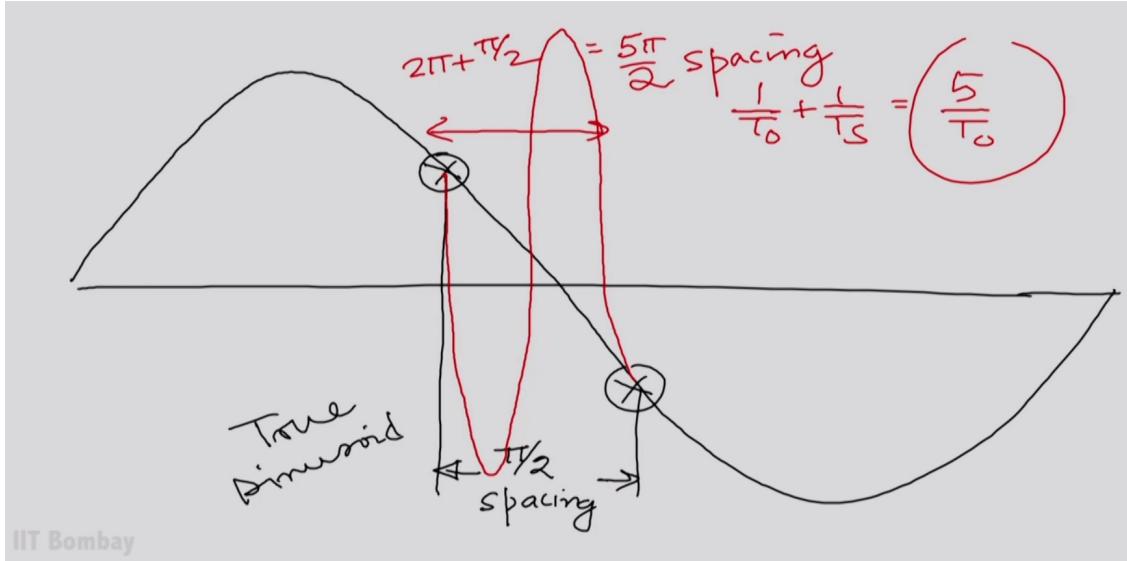


Figure 6.2.: True sinusoid and Ghost frequencies

Exercise : Draw the sinusoid when you almost miss the cycle and arrive on the wrong edge and verify the ghost frequency = $\frac{3}{T_0}$ and the spacing to be $\frac{3\pi}{2}$

6.3. Lecture 3: Concept and expression of Impostor Terms

6.3.1. Introduction

Previously, we have talked about “sampling” signals, that is, you only “read” and “store” the value of the signal at certain points. We saw the problems associated with sampling, mainly the fact that given just the sampled signal, it is not possible to reconstruct the original signal, because there are many possible signals which would have given the *same* samples, i.e. they would have taken on the same values as this signal at the points where we are sampling. The natural question to ask, then, is what “extra information” do we need in order to reconstruct the original signal from this sample? To do that, we first have to know what these “other signals” look like.

6.3.2. Non-Uniqueness of Samples

Previously, we saw graphically that many different sinusoids can have the same samples. We'll now do essentially the same thing, algebraically.

A sinusoid, in general, is of the form $A_0 \cos(\Omega_0 t + \phi_0)$, where $\Omega_0 = \frac{2\pi}{T}$, T being the time period, and ϕ_0 is the initial phase.

Now we saw that the process of sampling was also creating “ghost frequencies”, frequencies different from that of the original signal, as a side-effect. Of course, these “impostor terms”, as we can call them, have the same values as the original signal at the point of sampling. We also saw how you can lose a full cycle and arrive at the correct edge, or you only lose half a cycle and arrive at the wrong edge.

Let us represent these “monster frequencies” as

$$A_0 \cos(\Omega_{kj} t + \phi_{kj}) \quad (6.3)$$

Here, k is the number of cycles lost (including the half cycles, if any). $j = 1$ signifies that you arrived at the correct edge, and $j = 2$ stands for arriving at the wrong edge.

6.4. Lecture 4: Understanding the Impostor Terms

Now suppose we uniformly sample the signal at points $t = nT_s$. We have:

$$x(nT_s) = A_0 \cos(\Omega_0 nT_s + \phi_0)$$

Because of the periodicity of the cosine, we can add $\pm 2\pi nk$ to the phase without changing the value of the expression.

$$x(nT_s) = A_0 \cos(\Omega_0 nT_s \pm 2\pi nk + \phi_0) = A_0 \cos\left\{2\pi nT_s \left(\frac{1}{T_0} \pm \frac{k}{T_s}\right) + \phi_0\right\}$$

Here, one can think of the term $\frac{1}{T_0} \pm \frac{k}{T_s}$ as the frequency of the impostor in cycles/second, where $k \in \mathbb{Z}^+$.

Now let us assume, for the time being, that we are sampling the sinusoid at least one in each cycle. This is a very reasonable assumption, because if you take less than one sample per cycle, your original sinusoid is itself completing a whole cycle before the next sample is taken. Thus, we assume that $T_0 > T_s$. This would imply that $\frac{1}{T_0} - \frac{k}{T_s} < 0$, which says that the frequency is negative, which is something we don't like! But that's not too bad! We know that cos is symmetric about the y -axis. Hence, instead of only one expression with the \pm , we write two separate expressions:

$$A_0 \cos\left\{2\pi nT_s \left(\frac{1}{T_0} + \frac{k}{T_s}\right) + \phi_0\right\}$$

and

$$A_0 \cos\left\{-\left\{2\pi nT_s \left(\frac{1}{T_0} - \frac{k}{T_s}\right) + \phi_0\right\}\right\} = A_0 \cos\left\{2\pi nT_s \left(\frac{k}{T_s} - \frac{1}{T_0}\right) - \phi_0\right\}$$

Here, the first expression represents that we arrived on the correct edge after skipping k cycles completely, and the second one means that we arrived at the incorrect edge after “almost” skipping k cycles. Hence, with respect to the expression (6.3) which we wrote earlier, we can say that

$$\Omega_{k1} = 2\pi \left(\frac{k}{T_s} + \frac{1}{T_0}\right) \quad \phi_{k1} = \phi_0 \quad (6.4)$$

$$\Omega_{k2} = 2\pi \left(\frac{k}{T_s} - \frac{1}{T_0}\right) \quad \phi_{k2} = -\phi_0 \quad (6.5)$$

It is worthwhile to note that the signs of $\frac{1}{T_0}$ and ϕ_0 match in both expressions.

6.4. Lecture 4: Understanding the Impostor Terms

6.4.1. Introduction

We have talked about how sampling a signal can create impostor terms. As discussed previously, these impostor terms have the same values at the sampled points, but are in essence different signals all together. We also looked at the impostor terms for when the original signal is a sinusoid and the impostor terms are also sinusoids. In this lecture, we will carry this discussion forward and see how the different sinusoids act all together.

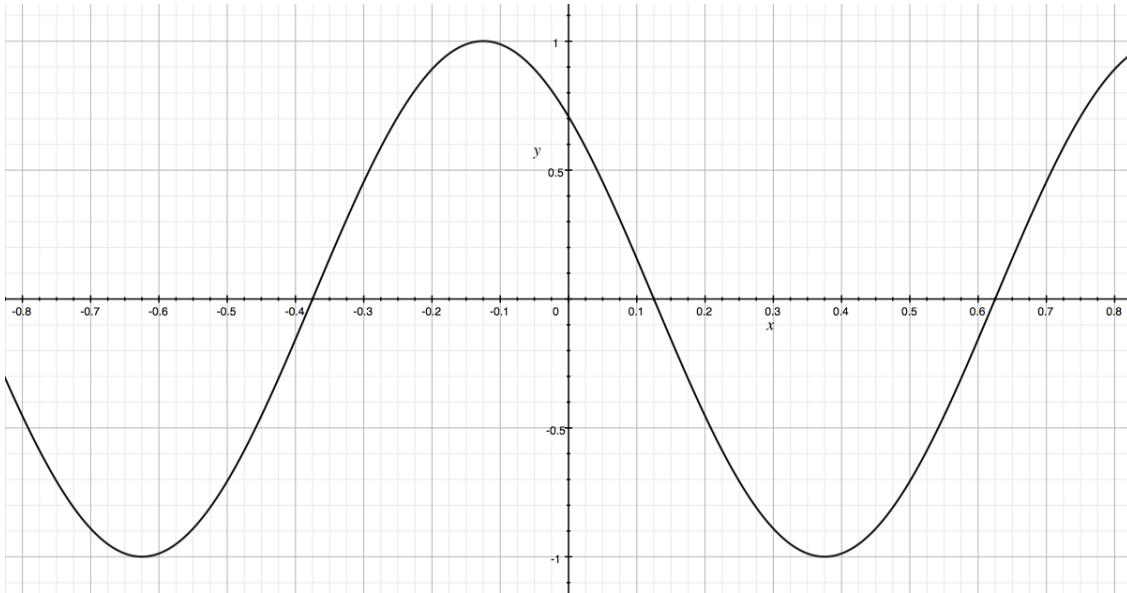


Figure 6.3.: The original signal, $y = \sin\left(\frac{2\pi t}{T_0} + \frac{\pi}{4}\right)$

6.4.2. Adding the sinusoid terms

Recall that the sinusoid terms can be written as follows:-

$$A_0 \cos\left(2\pi\left(\frac{k}{T_s} - \frac{1}{T_0}\right)t - \frac{\pi}{4}\right)$$

$$A_0 \cos\left(2\pi\left(\frac{k}{T_s} + \frac{1}{T_0}\right)t + \frac{\pi}{4}\right)$$

where k is any positive integer.

A sum of the original signal and all the impostor terms will look like:-

$$A_0 \cos\left(2\frac{\pi}{T_0}t + \frac{\pi}{4}\right) + \sum_{k=1}^N \left(A_0 \cos\left(2\pi\left(\frac{k}{T_s} - \frac{1}{T_0}\right)t - \frac{\pi}{4}\right) + A_0 \cos\left(2\pi\left(\frac{k}{T_s} + \frac{1}{T_0}\right)t + \frac{\pi}{4}\right) \right)$$

where $N > 0$

At this point, the following substitution will make it easier to work out the math.

$$B = \frac{2\pi k}{T_s} \text{ and } C = \frac{1}{T_0} + \frac{\pi}{4}$$

Notice that the term C is independent of k. For a given k, the sum inside the Σ reduces to

$$A_0 \cos(B - C) + A_0 \cos(B + C) = 2A_0 \cos B \cos C$$

This will make the sum :-

$$A_0 \cos\left(2\frac{\pi}{T_0}t + \frac{\pi}{4}\right) + \sum_{k=1}^N \left(2A_0 \cos\left(2\frac{\pi}{T_0}t + \frac{\pi}{4}\right) \cos\left(\frac{2\pi k}{T_s}\right) \right)$$

which can be further computed as:-

$$A_0 \cos(C) + \sum_{k=1}^N (2A_0 \cos(C) \cos(B))$$

6.4. Lecture 4: Understanding the Impostor Terms

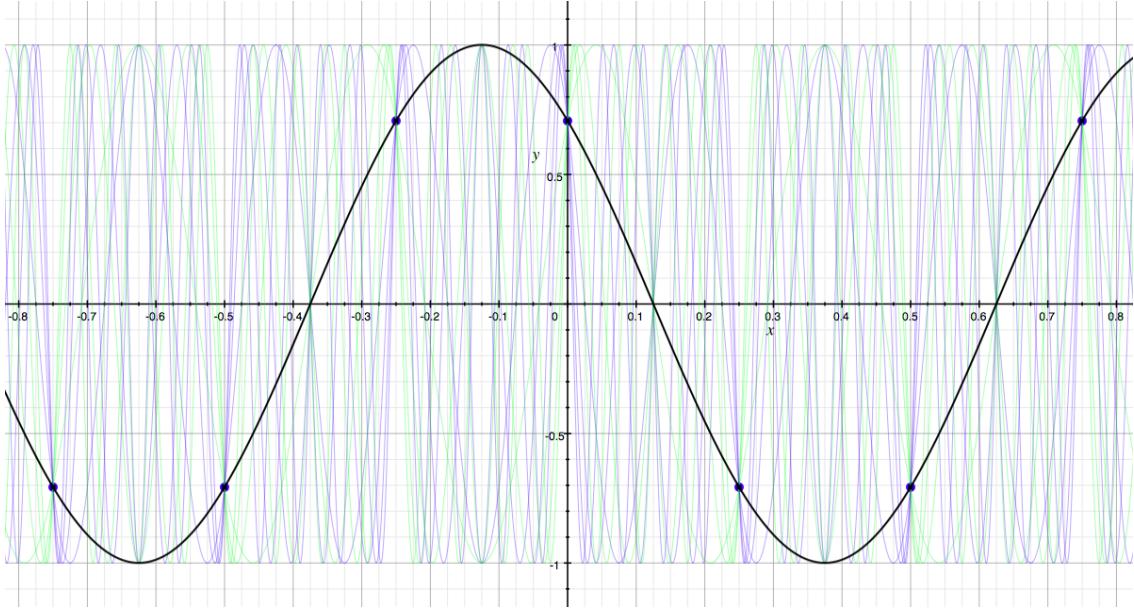


Figure 6.4.: The original signal, from Figure 6.3, sampled at points $t = nT_s, \forall n \in \mathbb{Z}$

$$A_0 \cos C \left\{ 1 + 2 \sum_{k=1}^N \cos B \right\}$$

Using the polar representation of $\cos \theta$, we can refine the sum in the brackets to be:-

$$\left\{ 1 + \sum_{k=1}^N \left(\exp\left(\frac{j2\pi kt}{T_s}\right) + \exp\left(-\frac{j2\pi kt}{T_s}\right) \right) \right\}$$

which can also be written as:-

$$1 + 2 \sum_{k=1}^{\infty} \cos\left(\frac{2\pi}{T_s} kt\right)$$

Keep this in mind because we will later draw a similar conclusion from a quite different manipulation. Note that the individual terms of the sum of exponentials can be added using the series for a Geometric progression quite easily. The first few steps are as follows:-

$$\begin{aligned} \sum_{k=1}^N \left(\exp\left(\frac{j2\pi kt}{T_s}\right) \right) &= \exp\left(\frac{j2\pi t}{T_s}\right) \left(\frac{1 - \exp\left(\frac{j2\pi t(N+1)}{T_s}\right)}{1 - \exp\left(\frac{j2\pi t}{T_s}\right)} \right) \\ &= \exp\left(\frac{j2\pi t}{T_s}\right) \left(\exp\left(\left(\frac{j2\pi t}{T_s}\right)\left(\frac{N+1}{2}\right)\right) \times \left(2j \sin\left(\left(\frac{j2\pi t}{T_s}\right)\left(\frac{N+1}{2}\right)\right) \right) \right. \\ &\quad \div \\ &\quad \left. \left(\exp\left(\left(\frac{j2\pi}{T_s}\right)\left(\frac{t}{2}\right)\right) \times \left(2j \sin\left(\left(\frac{j2\pi}{T_s}\right)\left(\frac{t}{2}\right)\right) \right) \right) \end{aligned}$$

The following are left as exercises to the reader:-

- Sketching this function as a function of t after further simplification and arithmetic
- Analysis of the function for when N grows.

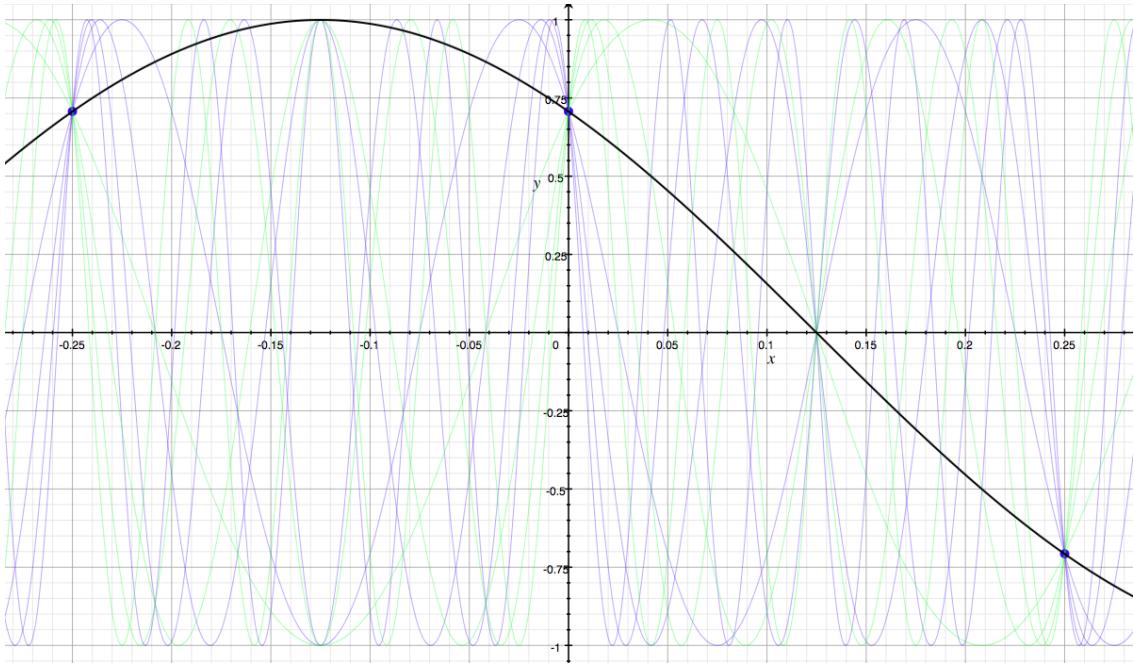


Figure 6.5.: Zoomed in view of the same signal

- Analysis of the function for when $t = nT_s$ and when $t \neq nT_s$
- Proceeding similarly for the other term in the original sum i.e. $\exp\left(\frac{-j2\pi kt}{T_s}\right)$ and conducting the same analysis for the resultant function.

6.4.3. An Alternate Path - By Sampling

The previous section contained a lot of mathematical computations with a large number of terms. We would like to tackle the problem in an alternate way which doesn't involve such heavy algebra.

Such an alternate way is to proceed by actually sampling the signal. By sampling the signal, we mean that the signal is retained around the sampling times (nt_s) for a time period of Δ ; i.e. the width of the signal is Δ .

The sampled curve can also be seen as the multiplication of the original signal (the sinusoid) with a train of very narrow pulses with unit height. However, note that these pulses are not impulses because they have a finite height (unit in this case).

Also note that this train of impulses is an even periodic function and hence, will have a Fourier series expansion of the form:-

$$c_0 + \sum_{k=1}^{\infty} c_k \cos\left(\frac{2\pi}{T_s} kt + 0\right)$$

The phase is 0 because the function is even. $k = 1$ represents the "fundamental" while $k > 1$ represents the "harmonics".

Using the previously learnt methods of computing the Fourier coefficients,

$$c_0 = \frac{1 \cdot \Delta}{T_s}$$

6.4. Lecture 4: Understanding the Impostor Terms

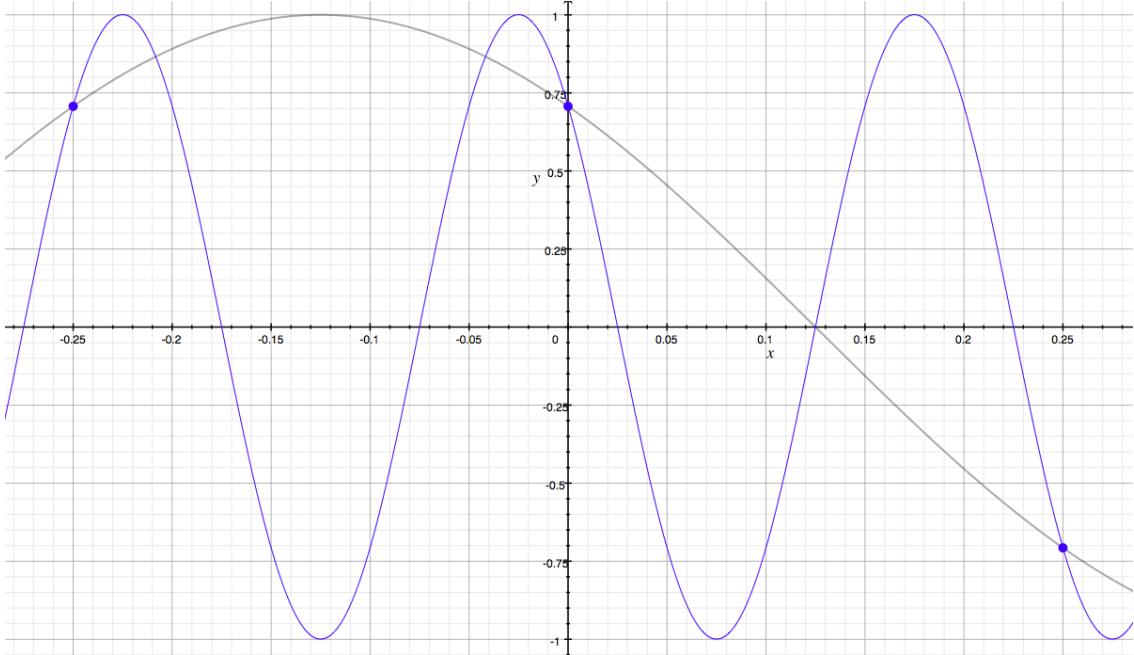


Figure 6.6.: Missing 1 cycle and arriving at the correct edge

and

$$c_k = \frac{2}{T_s} \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} 1 \cdot \cos\left(\frac{2\pi}{T_s} kt\right) dt$$

$$= \frac{2}{T_s} \frac{\sin\left(\frac{2\pi}{T_s} k\Delta\right)}{\left(\frac{2\pi}{T_s} k\right)} \Big|_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}}$$

On substituting the limits and simplifying the value,

$$c_k = \frac{2}{k} \sin\left(\frac{k\pi\Delta}{T_s}\right)$$

We see that the Fourier series becomes trivial for when $\Delta \rightarrow 0$. To prevent this, we employ the same technique as we did in Module 1, we modify the pulse in such a way that the height is the inverse of the width. Note that this had converted our *pulse* to an *impulse*.

The Fourier coefficients now become,

$$c_k = \frac{1}{T_s} \cdot \frac{2}{\frac{k\Delta}{T_s}} \sin\left(\frac{k\pi\Delta}{T_s}\right)$$

Now, the Fourier expansion doesn't become trivial when $\Delta \rightarrow 0$. Rather, we can see that this is nothing but the sinc function.

Hence, we can see that

$$c_k \rightarrow \frac{2\pi}{T_s} \quad \text{as} \quad \Delta \rightarrow 0$$

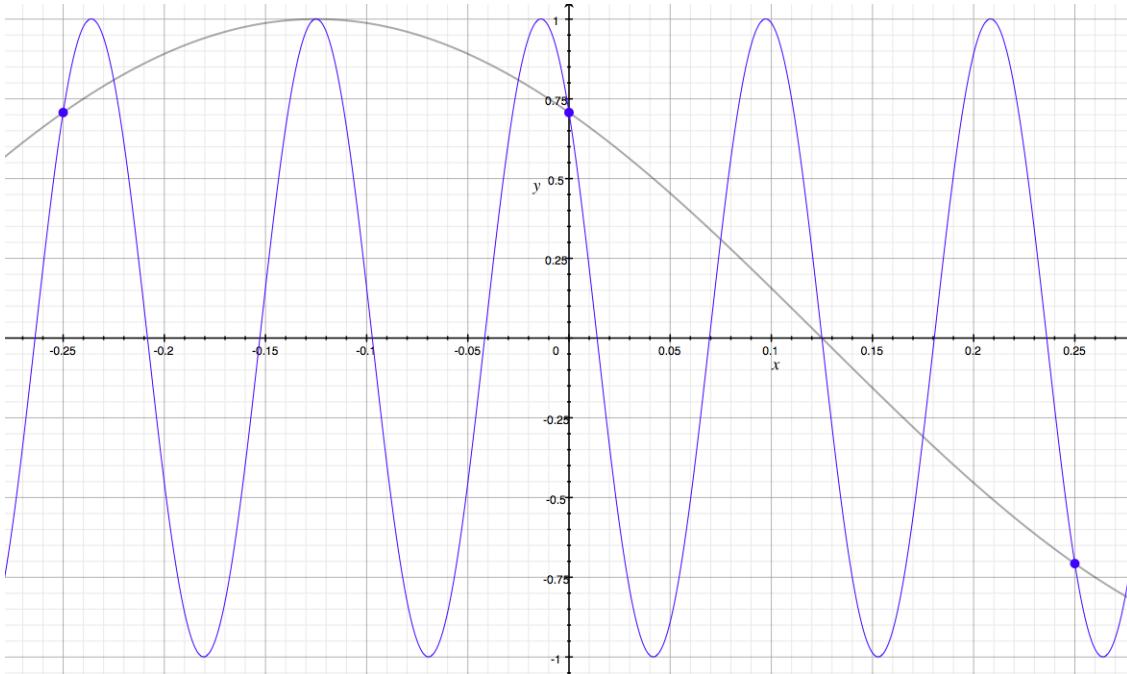


Figure 6.7.: Missing 2 cycles and arriving at the correct edge

6.4.4. Another Way of Sampling

We can also write the complex Fourier expansion of the train of impulses. And proceed to multiply the original signal by the Fourier expansion (complex expansion) of the train of impulses. This is different from our previous method in the aspect that we have started with a train of impulses straightaway; instead of considering pulses and then converting them to impulses. This will give us the same result.

The complex Fourier expansion of this train of impulses is as follows; where γ_l are the coefficients of each term of the expansion

$$f(t) = \sum_{l=-\infty}^{\infty} \gamma_l \cdot \exp(j \frac{2\pi}{T_s} lt)$$

where $f(t)$ is the train of impulses.

Let $p(t)$ denote any one of the impulses which are periodically repeated over the entire x -axis to get $f(t)$. Then for finding γ_l :-

$$\gamma_l = \frac{1}{T_s} \int_{-\frac{T_s}{2}}^{\frac{T_s}{2}} \left(p(t) \cdot \exp(-j \frac{2\pi}{T_s} lt) \right) dt$$

By the sifting property of the impulse, we can show that :-

$$\gamma_l = \frac{1}{T_s}$$

Hence, the complex Fourier expansion of the train of impulses is:-

$$f(t) = \frac{1}{T_s} \sum_{l=-\infty}^{\infty} \exp(j \frac{2\pi}{T_s} lt)$$

6.4. Lecture 4: Understanding the Impostor Terms

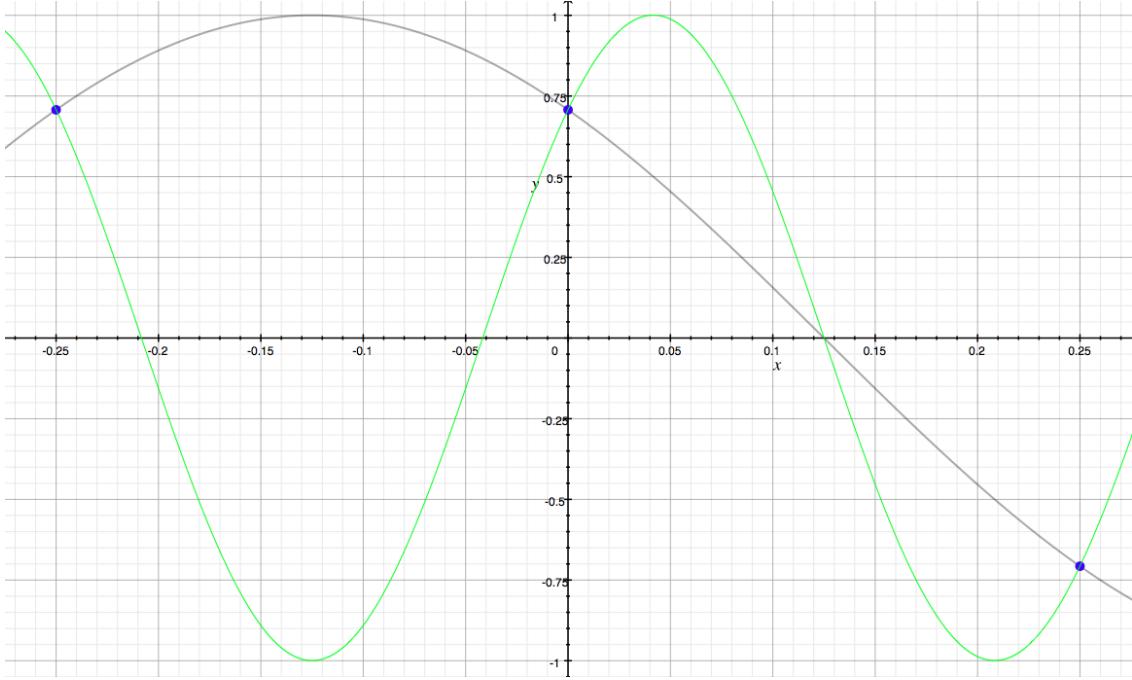


Figure 6.8.: Missing 1 cycle and arriving at the incorrect edge

which can further be simplified by taking the term for $l = 0$ out of the sum and clubbing the terms with the same value of l , but with different signs together:-

$$f(t) = \frac{1}{T_s} + \frac{1}{T_s} \sum_{l=1}^{\infty} \exp(j \frac{2\pi}{T_s} lt) + \exp(-j \frac{2\pi}{T_s} lt)$$

This can be further simplified as:-

$$f(t) = \frac{1}{T_s} + \frac{2}{T_s} \sum_{l=1}^{\infty} \cos\left(\frac{2\pi}{T_s} lt\right)$$

This is the same as what we saw when we added all the impostor terms together!

6.4.5. Conclusions

We took two paths to analysing the imposter terms and the original sinusoid; the first was adding all these terms while the second was sampling the signal using impulses and analysing the resultant signal. For the analysis with impulses, we took two ways; the real Fourier series and the complex Fourier series expansions of the train of impulses.

Our observation is that both these processes are in fact, the same!

So, to conclude, bringing all these "impostor" sinusoids together with the original amounts to multiplying the original sinusoid with a uniform train of impulses located at the sampling instants!

Another indirect observation is that since the addition of all the impostor sinusoids with the original is the same as the sampling of the original sinusoid, they have constructive interference at the sampling instants (and hence return a finite value) and have a destructive interference everywhere else (hence the curve vanishes everywhere else).

These conclusions should be kept in mind as they are essential to the development of concepts that will be introduced further.

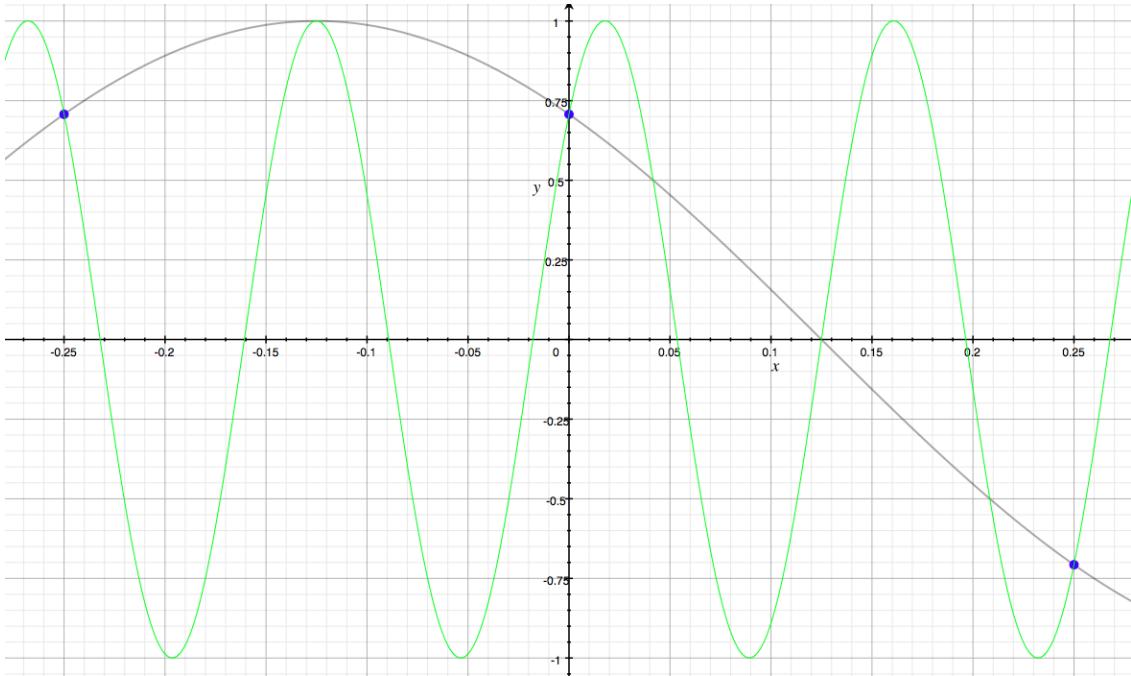
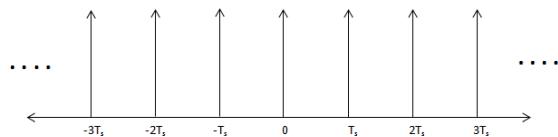


Figure 6.9.: Missing 2 cycles and arriving at the incorrect edge

6.5. Lecture 5: Sampling of signals having Fourier Transform

6.5.1. Sampling

Ideal sampling is multiplication of the waveform by a uniform train of impulses located at sampling instants. If $x(t)$ is being sampled, we are multiplying $x(t)$ with $p(t)$ which has impulses located at every sampling instants. $p(t)$ is shown below.



Altrnately, we are multiplying original function $x(t)$ by complex fourier expansion of the uniform impulse train.

Sampling of sinusoid

Let us focus on one constituent component, or sinusoid in the original waveform

$$A_0 \cos\left(\frac{2\pi t}{T_0} + \phi_0\right)$$

Now we are generalizing it to any phase ϕ_0 unlike we took $\frac{\pi}{4}$ in previous lecture. (Any ϕ_0 will do)

Ideal sampling results in this becoming,

$$K_0 \left\{ A_0 \cos\left(\frac{2\pi t}{T_0} + \phi_0\right) + \sum_{l=1}^{\infty} \left\{ A_0 \cos\left[2\pi\left(\frac{l}{T_s} - \frac{1}{T_0}\right)t - \phi_0\right] + A_0 \cos\left[2\pi\left(\frac{l}{T_s} + \frac{1}{T_0}\right)t + \phi_0\right]\right\} \right\}$$

6.5. Lecture 5: Sampling of signals having Fourier Transform

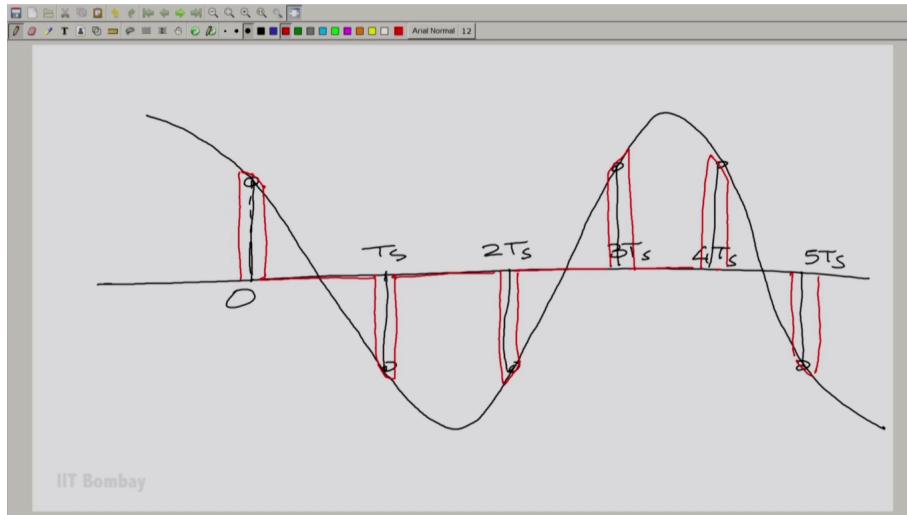


Figure 6.10.: The red curve represents the sampled signal while the black curve represents the original signal

where K_0 is some constant depends on the strength of the impulses. We have to make sure that all these frequencies are distinct i.e, the imposters do not overlap on original frequencies. As long as $\frac{1}{T_0}$ much less than $\frac{1}{T_s}$, all these are distinct frequencies.

Example: $\frac{1}{T_0} = 1$ khz, $\frac{1}{T_s} = 20$ khz

The frequencies are

$$\frac{1}{T_0}, \left(\frac{1}{T_s} - \frac{1}{T_0}\right), \left(\frac{1}{T_s} + \frac{1}{T_0}\right), \left(\frac{2}{T_s} - \frac{1}{T_0}\right), \left(\frac{2}{T_s} + \frac{1}{T_0}\right) \dots$$

The frequencies are 1, 19, 21, 39, 41, Clearly they are distinct!

How small can $\frac{1}{T_s}$ be to keep the frequencies distinct ?

Keeping imposters distinct

We are keen on the frequencies being distinct because imposters should be distinguishable from original waveform.

$$\left(\frac{1}{T_s} - \frac{1}{T_0}\right) < \left(\frac{1}{T_s} + \frac{1}{T_0}\right)$$

We are doubtful if $\frac{1}{T_0}$ and $\frac{1}{T_s} - \frac{1}{T_0}$ are distinct. They are distinct if

$$\left(\frac{1}{T_s} - \frac{1}{T_0}\right) > \frac{1}{T_0}$$

$$\frac{1}{T_s} > \frac{2}{T_0}$$

So, any sampling rate $\frac{1}{T_s}$ greater than (STRICTLY GREATER) twice of original frequency $\frac{1}{T_0}$. That means $\frac{1}{T_s} = 3$ khz is enough. No need of taking samples at 20 khz frequency. For sampling rate $\frac{1}{T_s}$ the frequencies are 1, 2, 4, 5, 7, 8, 10. Notice they are still distinct.

Also note

$$\frac{1}{T_s} \neq 2 \text{ khz}$$

as 1st imposter sits on original waveform.

6.5.2. Signals which have and do not have a Fourier transform

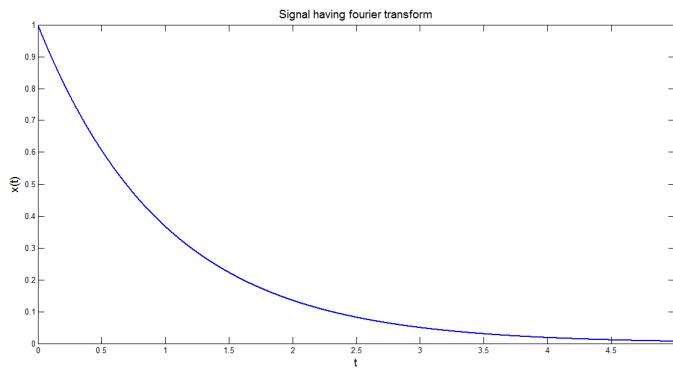
A waveform or signal is comprised of many sinusoids, the FT of the signal may contain discrete frequencies (case where $x(t)$ is periodic), or there can be continuum of frequencies (when $x(t)$ is aperiodic).

When we say a waveform is composed of many constituents, we say that waveform has fourier transform.

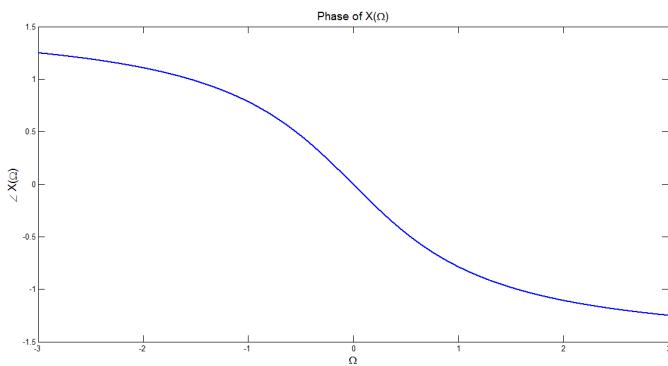
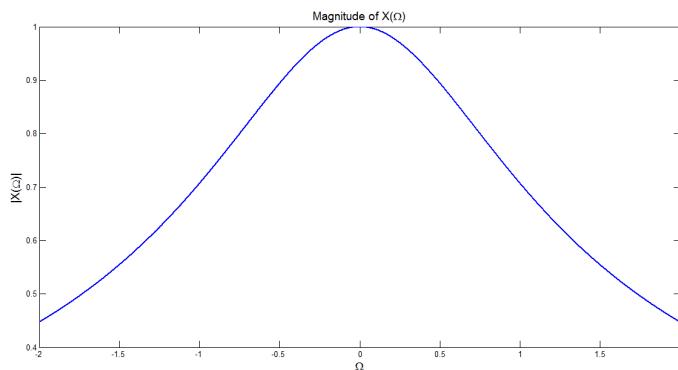
Signal having Fourier Transform

An example of signal that has fourier transform :-

$$x(t) = e^{-t}u(t), u(t) = \text{standard unit step}$$



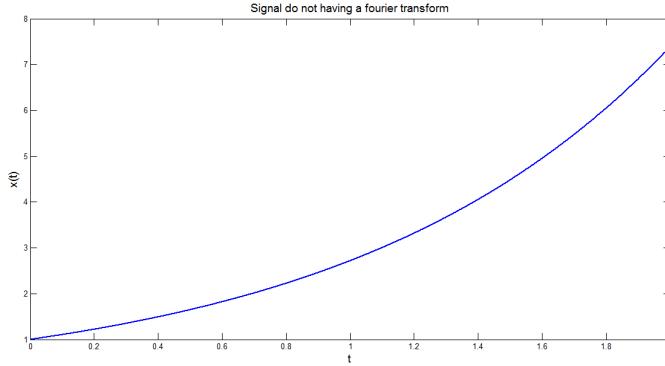
$$X(\Omega) = \int_{-\infty}^{+\infty} x(t)e^{-j\Omega t} dt = \int_{-\infty}^{+\infty} e^{-t}e^{-j\Omega t} dt = \int_{-\infty}^{+\infty} e^{-(1+j\Omega t)} dt = \frac{1}{1+j\Omega}$$



Signals not having Fourier Transform

An example for signal which does not have a fourier transform:- $y(t) = e^t u(t)$.

$$X(\Omega) = \int_{-\infty}^{+\infty} y(t) e^{-j\Omega t} dt = \int_{-\infty}^{+\infty} e^t e^{-j\Omega t} dt$$



Clearly the integral does not converge. This signal is divergent and so no fourier transform. Throughout the discussion on sampling we omit signals which do not have fourier transform from consideration. We only focus on signal having Fourier transform.

$$x(t) = e^{-t} u(t) \iff X(\Omega) = \frac{1}{1 + j\Omega}$$

We can reconstruct $x(t)$ from $x(\Omega)$

$$x(t) = \int_{-\infty}^{+\infty} X(\Omega) e^{j\Omega t} d\Omega$$

Recall in the geometric interpretation $X(\Omega)$ is component, $e^{j\Omega t}$ is vector, $d\Omega$ is aggregated over all vector and $\frac{1}{2\pi}$ normalizes.

$$\begin{aligned} X(\Omega) &= \overline{X(-\Omega)} \\ |X(\Omega)| &= |X(-\Omega)| \\ \angle X(\Omega) &= -\angle X(-\Omega) \end{aligned}$$

Consider $X(\Omega)e^{j\Omega t} + X(-\Omega)e^{-j\Omega t}$, as one of the constituent. On expansion

$$|X(\Omega)|e^{j\angle X(\Omega)}e^{j\Omega t} + |X(-\Omega)|e^{j\angle X(-\Omega)}e^{-j\Omega t} = 2|x(\Omega)|\{\cos(\Omega t + \angle x(\Omega))\}d\Omega$$

We have continuum of such constituents (Ω from 0 to ∞).

$x(t)$ is combination of $2|X(\Omega)|\{\cos(\Omega t + \angle x(\Omega))\}$ for all Ω .

$$x(t) = \int_0^{\infty} \frac{2}{\sqrt{1 + \Omega^2}} \cos(\Omega t - \tan^{-1} \Omega) d\Omega$$

Note that for every $\Omega \neq 0$, we designed our constituent as combination of Ω and $-\Omega$ terms. Hence limits vary from 0 to ∞ . Now, when we sample $x(t)$ each and every such constituent is sampled and creates its own imposters. Can we keep all those imposters distinct from original frequencies? Can we ever sample "adequately" so that we can distinguish imposters from original frequencies?

6.5.3. Linearity of sampling

Recall constituent frequency $\frac{1}{T_0}$, sampling frequency $\frac{1}{T_s}$. Frequencies created in sampling are

$$\frac{1}{T_0}, \frac{1}{T_s} - \frac{1}{T_0}, \frac{1}{T_s} + \frac{1}{T_0}, \frac{2}{T_s} - \frac{1}{T_0}, \frac{2}{T_s} + \frac{1}{T_0}, \dots$$

The smallest $\frac{1}{T_s}$ we could use is $> \frac{2}{T_0}$ where $\frac{1}{T_0}$ was frequency of one constituent.

When we have combination of sinusoids, can we generalize constituent role ? Here a basic question arises: IS SAMPLING LINEAR?

Yes. Sampling ideally means multiplying the signal by a uniform train of impulse. sampling of $x(t)$ is $x(t)p(t)$

$$x_1(t) \Rightarrow x_1(t)p(t)$$

sampled at rate $\frac{1}{T_s}$

$$x_2(t) \Rightarrow x_2(t)p(t)$$

sampled at rate $\frac{1}{T_s}$ Now consider a linear combination of $x_1(t)$ and $x_2(t)$

$$\alpha x_1(t) + \beta x_2(t) \Rightarrow \{\alpha x_1(t) + \beta x_2(t)\}p(t) = \alpha\{x_1(t)p(t)\} + \beta\{x_2(t)p(t)\}$$

Sampling a linear combination is equivalent to the some linear combination of samples.

Sampling any one of the constituents would create imposters $\frac{1}{T_0}, \frac{1}{T_s} - \frac{1}{T_0}, \frac{1}{T_s} + \frac{1}{T_0}, \frac{2}{T_s} - \frac{1}{T_0}, \frac{2}{T_s} + \frac{1}{T_0}, \dots$ For every T_0 it will happen.

Effect of sampling a signal with multiple constituents \cong sum of effects of sampling each constituents.

Example : Let us have two constituent frequencies in the original waveform $\frac{1}{T_{01}}, \frac{1}{T_{02}}$ @ sample rate $= \frac{1}{T_s}$

The imposters are as follows

$$\frac{1}{T_s} - \frac{1}{T_{01}}, \frac{1}{T_s} - \frac{1}{T_{02}}, \frac{1}{T_s} + \frac{1}{T_{01}}, \frac{1}{T_s} + \frac{1}{T_{02}}, \frac{2}{T_s} - \frac{1}{T_{01}}, \frac{2}{T_s} - \frac{1}{T_{02}}, \frac{2}{T_s} + \frac{1}{T_{01}}, \frac{2}{T_s} + \frac{1}{T_{02}}, \dots$$

Is it possible to keep imposters distinguished from constituents for all such constituents? Obviously the answer in this case is NO.

As our sampling rate is finite, every constituent frequency in your signal needs to be finite otherwise we can never distinguish imposters! We need to limit the constituent frequencies!! There must be an upper limit on constituent frequency which is nothing but BAND LIMITED SIGNAL. Sampling must be “adequate” or take care of imposter.

If our signal is BAND LIMITED what minimum sampling rate do you think would keep those imposters distinct ? Let us answer this in next lecture .

6.6. Lecture 6: Sampling Theorem

6.6.1. Introduction

Sampling is broadly defined as the process of taking samples of an analog signal so as to convert it to a digital sequence. Sampling is the fundamental process that converts analog signal to digital signal.

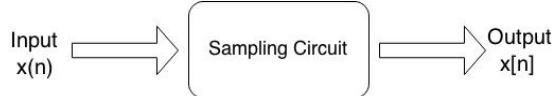


Figure 6.11.: Block Diagram of a Sampling System.

One of the best examples of application of sampling could be that of an Analog to Digital Converter (ADC). An ADC continuously take samples of the incoming data and gives an output as a sequence of discrete signals.

The audio CDs that we play on our music systems are recorded by continuous sampling of the incoming audio signal (music) and are written on to the CD as a sequence of Digital Data sequence.

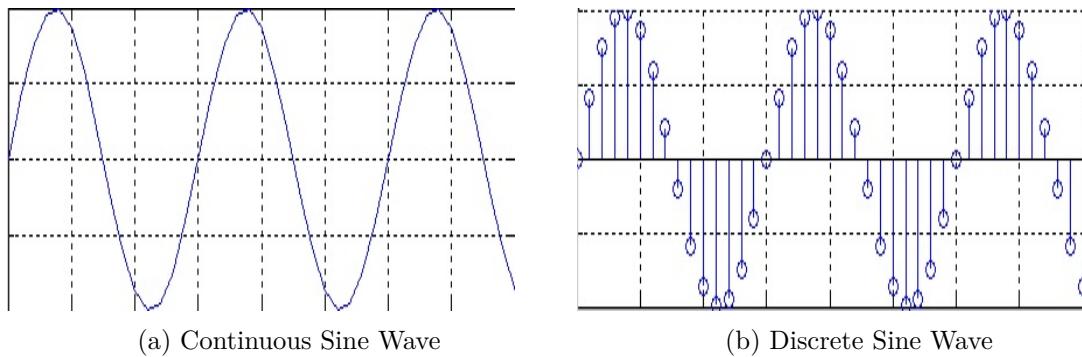


Figure 6.12.: Sampling of a Continuous Sine Wave

6.6.2. Sampling Theorem

Also known as "Shannon-Whittaker-Nyquist theorem on Sampling".

When a signal is sampled, there is a possibility to lose information. During the sampling process, a lot of unwanted information maybe clubbed with the actual useful data. There is a need to filter out the unwanted part or the 'imposters' created by sampling.

Considerations for a given signal $x(t)$:

1. The signal has a Fourier Transform.
2. The Fourier Transform is non-zero up to a maximum finite frequency ' f_m '.

The sampled signal is sampled at a frequency ' f_s '.

The Sampling Theorem can be stated as:

A band-limited signal, band-limited to a maximum frequency f_m can be perfectly reconstructed from it's samples taken at a rate $f_s = \frac{1}{T_s}$, such that $f_s > 2f_m$.

6.6.3. Proof

Sampling is a linear process.

Net consequence of a sampling spectrum is given by the sum of contributions by individual spectral pieces.

Consider a signal $x(t)$, with Fourier Transform $X(f)$.

The signal will be sampled at a frequency at a frequency f_s .

6. chapter

The Nyquist sampling theorem states that for the complete reconstruction of a signal, the sampling frequency $f_s > 2f_m$.

If $f_s < 2f_m$, the 'imposters' created by sampling will cloud with the original signal. Thus, resulting in loss of data.

$f_s > 2f_m$ implies $f_s - f_m > f_m$.

The imposter frequency varies from f_s to $f_s - f_m$.

The Nyquist criteria ensure that since $f_s - f_m > f_m$, the imposter frequency does not coincide with the actual data.

Thus, reconstructing the entire signal is basically filtering out all the 'imposters' created and only obtain the original signal.

The process of sampling will be further discussed in the next session.

6.7. Lecture 7: Reconstructor: Ideal and practical

6.7.1. Introduction

In the previous lecture, we saw that a band-limited signal can be perfectly reconstructed from its samples, if it has been sampled at a frequency greater than twice the maximum frequency of the original signal f_m . In this lecture we will see what the procedure for reconstruction is by analysing in the frequency domain.

6.7.2. Frequency response for Ideal filter for reconstruction

It was shown that sampling creates multiple copies of the original spectrum shifted by integer multiples of the sampling frequency f_s . In order to fully reconstruct the original signal, we need a system which retains the original spectrum and cuts-off the copies as shown in Fig.6.13.

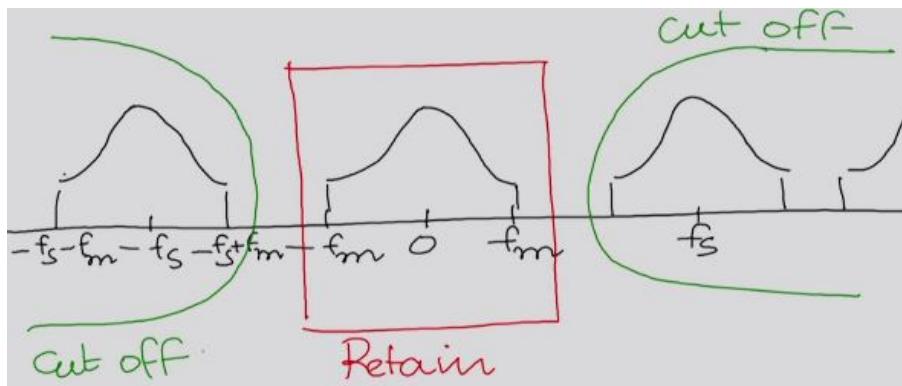


Figure 6.13.: Principle of reconstruction

We can achieve this by passing the sampled signal to a LSI system with a frequency response $H(f)$ as shown in Fig.6.14. The filter needs to leave the signal unchanged in the frequency range $-f_m - \delta$ to $f_m + \delta$ and cut off the rest of the frequencies. In order to ensure that the copies are completely outside this window, we require $f_m + \delta < f_s - f_m$. This is because from Fig.6.13, it can be seen that there is a margin between f_m and $f_s - f_m$. As long as the edge of the window falls inside this margin, we are good to go.

But why do we require this margin? Why do we need to sample at *greater* than twice f_m ? This is because it would create a problem if our signal had a pure sinusoid exactly at

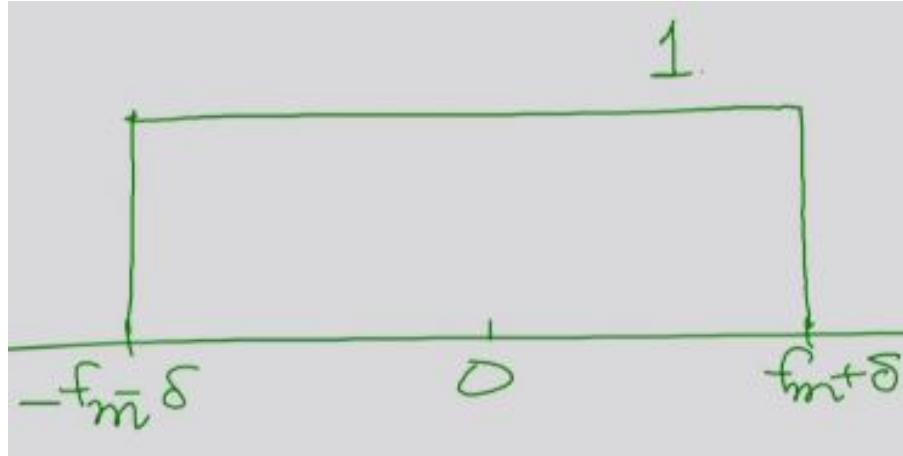
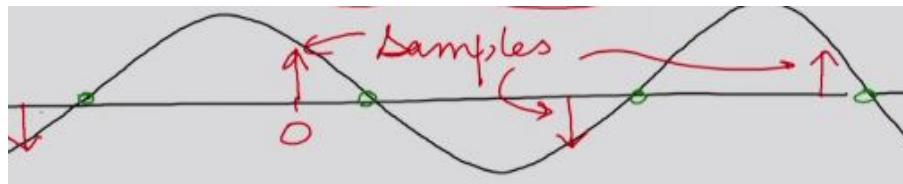


Figure 6.14.: Desired frequency response of filter

f_m i.e a *tonal* component. A pure sinusoid of frequency f_m in the time domain (a tonal component at f_m) would imply an impulse at $f = f_m$ in the frequency domain. Let us consider such a tonal component and see what happens if we were to sample it exactly at twice its frequency, $f_s = 2f_m$. This means that we have two samples in every period. In the unlucky event that we sampled at the zero-crossing of the sinusoid, all our samples will simply have a value 0 as shown in green in Fig.6.15.


 Figure 6.15.: Sampling at exactly twice f_m may yield all samples to be zero

Additionally, the filter should have a cut-off frequency beyond f_m as we want to retain the tonal component at f_m . Hence the margin is necessary.

6.7.3. Impulse response of ideal filter for reconstruction

To find this, we have to compute the inverse Fourier transform of the filter shown in Fig.6.14.

$$\begin{aligned}
 h(t) &= \int_{-\infty}^{\infty} H(f) e^{j2\pi ft} df \\
 &= \int_{-f_m-\delta}^{f_m+\delta} 1 \cdot e^{j2\pi ft} df \\
 &= \frac{e^{j2\pi ft}}{j2\pi t} \Big|_{-f_m-\delta}^{f_m+\delta} \\
 &= \frac{e^{j2\pi ft}}{j2\pi t} \Big|_{-f_c}^{f_c} \text{ where } f_c = f_m + \delta \\
 &= \frac{e^{j2\pi f_c t} - e^{-j2\pi f_c t}}{j2\pi t} \\
 &= \frac{2j \sin(2\pi f_c t)}{j2\pi t}
 \end{aligned} \tag{6.6}$$

The impulse response is basically a *sinc* function. It goes to zero at all integer values of the argument, i.e when $2f_c t$ is an integer.

6.7.4. Problems with the ideal reconstructor

Stability

Is the system given by the impulse response $h(t)$ as in (6.6) a stable system? It can be proven that it is not. For stability we require,

$$\int_{-\infty}^{\infty} |h(t)| dt < \infty \tag{6.7}$$

Challenge : Show that the above inequality does not hold. In other words, show that the absolute integral of $h(t)$ diverges thereby rendering the system unstable.

Hint : Split the integral into pieces. Find some quantity which is smaller than these pieces. Then show that the sum of the smaller quantity diverges meaning that the sum of the actual pieces will definitely diverge. $h(t)$ looks as in Fig.6.16. In $|h(t)|$, the negative

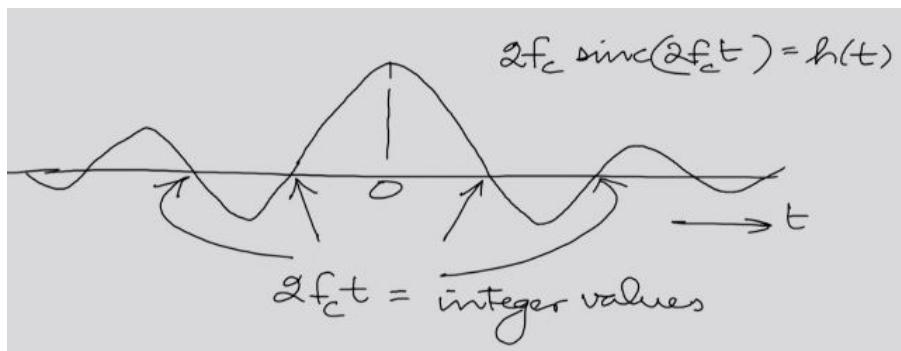


Figure 6.16.: Impulse response $h(t)$

parts get reflected about the x -axis and become positive.

Step 1 : Take each of the segments or lobes.

Step 2 : Lower bound each of these areas.

Step 3 : Show that the sum of these lower-bound areas is itself divergent.

Note : This kind of a frequency response is called as a *brick wall* response because of its structure. Brick-wall responses are always unstable systems. However although this system is unstable, its impulse response has a Fourier transform. In general this need not be the case. It means that if we give this system an input of a sinusoid, we are guaranteed a sinusoid output but there may very well be some bounded inputs whose outputs turn out to be unbounded. Hence while trying to reconstruct a bounded sampled signal, we may end up with an unbounded output if we were to use this filter.

What can we do about it? Recall the margin in the frequency response that we asked for earlier. We can make use of this to create a filter which is not brick wall-like and stable. Consider the frequency response of sampled signal (only with frequency values shown) as in Fig. 6.17. As there is no frequency present in the margin, we do not really care

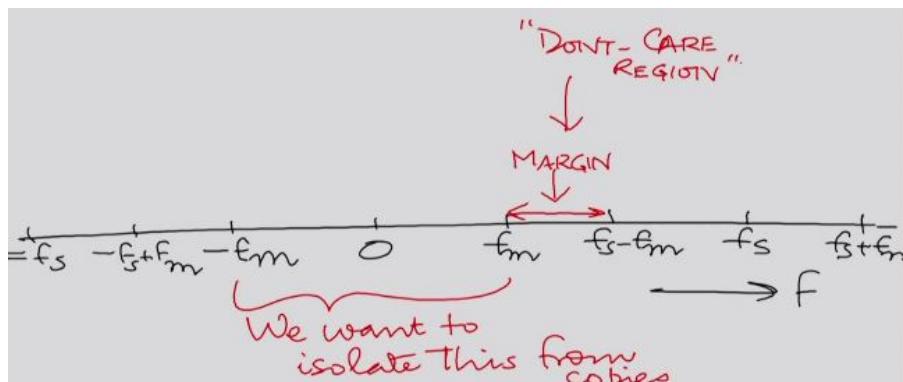


Figure 6.17.: Making use of the margin or the *don't care region* to construct a stable filter

what the filter does to those frequencies, or in other words, the frequency response at those frequencies is immaterial. Using this fact, we can use a filter which is not brickwall-like as in Fig.6.18, for example. This kind of a filter has a frequency response which

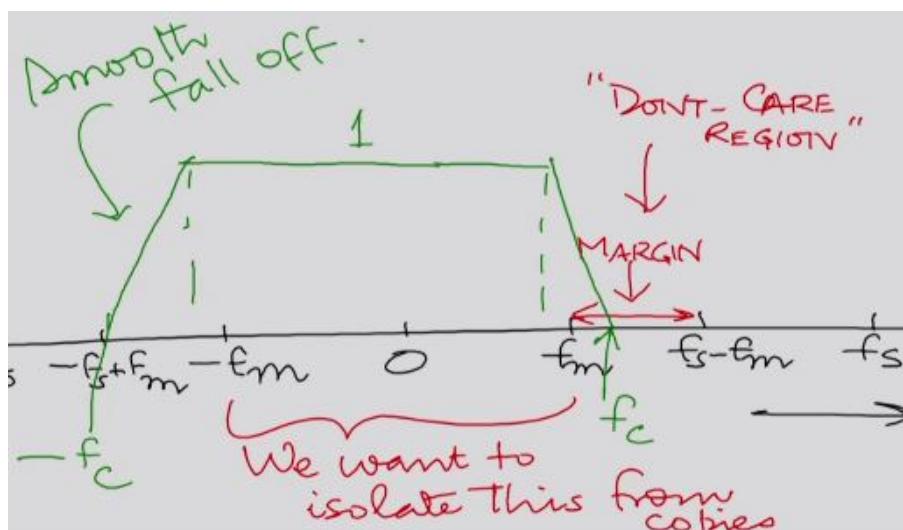


Figure 6.18.: Example of a filter which is not brickwall-like, but still serves our purpose

is *continuous*, unlike the brickwall filter (which had a discontinuity at $f = f_c$). So the margin has allowed us to make a stable filter. How large can the margin be? It is governed

6. chapter

by the difference between f_s and $2f_m$.

Realisability

Look at both the brickwall filter and the modified stable filter which was continuous. Both have a flat region between $-f_m$ and f_m . It is impossible to realise this practically. The flat region in the frequency response brings in *irrationality* into the response and irrational systems require infinite resources to realise.

We can make things more practical. Let us try to make a filter which varies *slowly* in the previously flat region. Beyond this region it can be allowed to vary fast (Fig. 6.19).

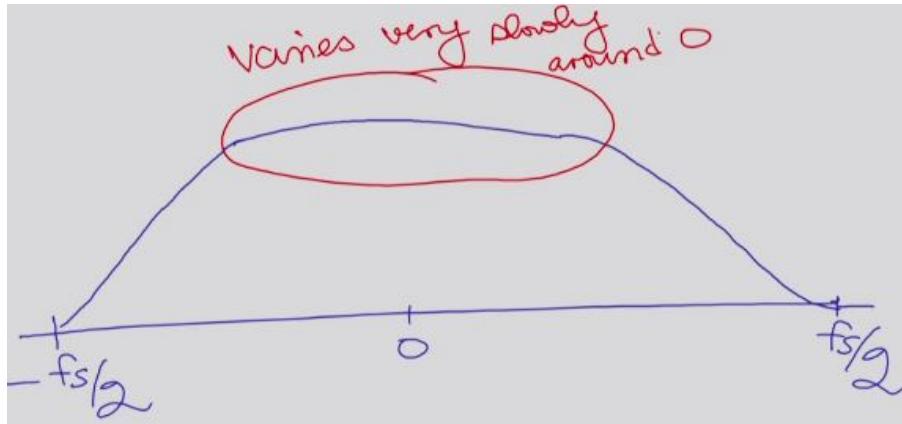


Figure 6.19.: A filter which is practically realisable

6.8. Lecture 8: An example of reconstructor

6.8.1. Introduction

In the previous lecture, we discussed about a practical reconstructor and its various properties. In this lecture, we discuss about a simple RC circuit which behaves as a reconstructor. We also discuss the magnitude response of an RC circuit, plot its graph and the limitations of ideal sampling.

6.8.2. An Example: RC circuit

In this lecture, the concept of sampling and reconstruction is made practical and physically realizable. The simple RC combination is used as a reconstructor. The following figure depicts an RC circuit:

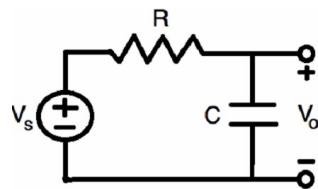


Figure 6.20.: RC circuit.

Graphical approach to calculate rate of change of frequency response

The frequency response of the given circuit is $H(\Omega)$.

Let Ω be the angular frequency.

Let $\tau = RC$.

Then,

$$H(\Omega) = \frac{1}{1+jRC\Omega}$$

Magnitude of frequency response of the RC circuit is

$$|H(\Omega)| = \left[\frac{1}{(1+(\tau\Omega)^2)} \right]^{1/2}$$

Therefore the graph of the magnitude response of the RC circuit is as shown below:

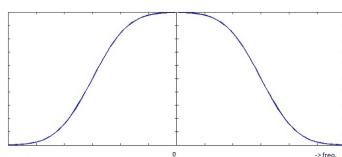


Figure 6.21.: Magnitude response of RC circuit

From the graph it is clear that the rate of magnitude change is less near 0, then increases for some time and then the rate of magnitude change again reduces and tends to 0.

Analytic approach to calculate rate of change of frequency response

Now we try to algebraically obtain the rate of change of magnitudes at various instants.

Taking derivative of $H(\Omega)$,

$$\begin{aligned} \frac{d}{d\Omega} |H(\Omega)| &= -\frac{1}{2} [1 + (\Omega\tau)^2]^{-3/2} \cdot \frac{d}{d\Omega} [1 + (\Omega\tau)^2] \\ &= \frac{-\tau^2\Omega}{[1 + (\Omega\tau)^2]^{3/2}} \end{aligned}$$

Case 1:

Now from the equation we can see that as

$\Omega \rightarrow 0$,

$$\frac{d}{d\Omega} |H(\Omega)| \rightarrow 0.$$

Case 2:

As $\Omega \rightarrow \infty$,

$$\frac{d}{d\Omega} H(\Omega) \approx \frac{\Omega}{\Omega^3} \approx \frac{1}{\Omega^2}$$

$$\text{Hence, } \frac{d}{d\Omega} H(\Omega) \rightarrow 0$$

Hence the rate of frequency response is very less near 0 and also when the values are very large, as indicated by the graph.

6.8.3. Properties of a realizable reconstructor

RC circuit is a simple and effective reconstruct. A stable and physically realizable reconstructor must NOT have:

1. A completely flat region.
2. A region of brickwall.
3. Completely zero magnitude region.

For a perfect reconstructor the response must become 0 after a certain interval, effectively after $f_s/2$. This is not possible for a RC circuit as it does have a non zero response after $f_s/2$. As there are no completely flat surfaces the following errors occur:

1. As the frequency response has no flat region, the reconstructed signal will have slight distortions.
2. As the reconstructor does not have zero response, the carbon copies also have an effect in the reconstructed signal.

To reduce these problems of ideal sampling, non-ideal sampling is used. The next session discusses about these practical non-ideal sampling.

6.9. Lecture 9: Nonideal Sampling with Pulses

6.9.1. Introduction

We are, in the current string of lectures, trying to make sense of the more practical aspects of sampling and reconstruction. We have seen so far that an RC circuit can be considered a very crude reconstructor, for a reasonable amount of scenarios. There are two conditions which we saw which must be satisfied in order for the RC circuit to work as reconstructor:

1. The signal should fall within the flat zone of the response of RC circuit.
2. We need to choose/modify our sampling process such that as $\Omega \rightarrow 0$, the nonzero fall-off of the signal doesn't allow too much from the carbon copies of the Fourier transform of the signal to come into output.

We will try to look how to go about satisfying the second condition:

6.9.2. The Pulse Train

We can satisfy the above condition by using nonideal sampling. So we will try to give a proof for a more generalized version of Shannon theorem which we have already seen. We will **not** be sampling with an impulse train but with a pulse train, as shown.

Let's call this pulse train ' $p(t)$ ' and let's try to find it's Fourier series expansion. We will use complex Fourier series, to write

$$C_k = \frac{1}{T_s} \int_0^{T_s} p(t) e^{-j \frac{2\pi}{T_s} kt} dt$$

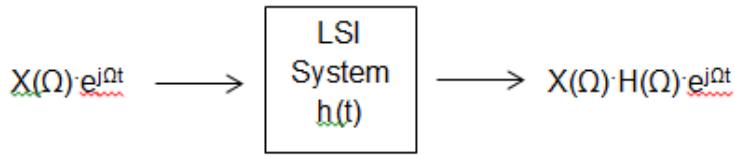


Figure 6.22.: The Pulse Train with period T_s

Where C_k is the coefficient of the k^{th} term in the Fourier series expansion, T_s is the sampling period, and $j = \sqrt{-1}$. So we can write that expression (for nonzero k) as

$$\begin{aligned}
C_k &= \frac{1}{T_s} \int_0^{T_s} p(t) e^{-j\frac{2\pi}{T_s}kt} dt \\
&= \frac{1}{T_s} \int_0^{\Delta} e^{-j\frac{2\pi}{T_s}kt} dt \quad \text{as } p(t) \text{ is nonzero, equal to one only in that region} \\
&= \frac{1}{T_s} \left(\frac{e^{-j\frac{2\pi}{T_s}k\Delta}}{-j\frac{2\pi}{T_s}k} \right)_0 \\
&= \frac{e^{-j\frac{2\pi}{T_s}k\Delta} - 1}{T_s \left(-j\frac{2\pi}{T_s}k \right)} \\
&= \frac{1 - e^{-j\frac{2\pi}{T_s}k\Delta}}{j2\pi k}
\end{aligned}$$

We can simplify this as

$$\begin{aligned}
C_k &= e^{-j\frac{2\pi}{T_s}k\frac{\Delta}{2}} \cdot \frac{2j \sin(\frac{2\pi}{T_s}k\frac{\Delta}{2})}{2j\pi k\frac{\Delta}{T_s}} \cdot \frac{\Delta}{T_s} \\
&= \frac{\Delta}{T_s} e^{-j\frac{2\pi}{T_s}k\frac{\Delta}{2}} \cdot \frac{\sin(\frac{\pi k\Delta}{T_s})}{\frac{\pi k\Delta}{T_s}} \\
&= \frac{\Delta}{T_s} e^{-j\frac{2\pi}{T_s}k\frac{\Delta}{2}} \cdot \text{sinc}\left(\frac{k\Delta}{T_s}\right)
\end{aligned}$$

Where sinc function is given by

$$\text{sinc}(r) = \frac{\sin(\pi r)}{\pi r}$$

We can also scale these pulses so as to have unit area, by multiplying by $\frac{1}{\Delta}$, and it would just cancel out the delta in our expression, so we'd have

$$C_k = \frac{1}{T_s} e^{-j\frac{2\pi}{T_s}k\frac{\Delta}{2}} \cdot \text{sinc}\left(\frac{k\Delta}{T_s}\right)$$

When $k = 0$, we can show that C_0 is the limit of this expression as $k \rightarrow 0$. We can also compute C_0 independently.

Let us plot the sinc function and see how it looks like:

In ideal sampling, we had $\frac{\Delta}{T_s}$ very small, tending to zero. In that case we had all

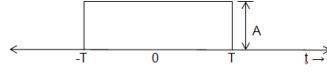


Figure 6.23.: The 'sinc' function.

our samples taken very closely, around zero. Applying $\Delta \rightarrow 0$ to the expression of C_k (scaled), we have $C_k \approx \frac{1}{T_s}$ i.e. all the C_k 's are same. However, if $\Delta \neq 0$, we have the expression for C_k as stated earlier, involving sinc function. We can write it as

$$C_k = \gamma_0 \text{sinc} \left(\frac{k\Delta}{T_s} \right)$$

Here γ_0 is a constant. Let's look at the Fourier series representation of $p(t)$: Note

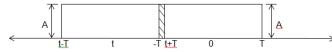


Figure 6.24.: The sampled version of $\text{sinc}(r)$. r is on the x-axis. Some of the C_i 's are labeled.

that the distance between two consecutive samples is $\frac{\Delta}{T_s}$. We can see that as $\Delta \rightarrow 0$, we will have all our samples around zero. If $\frac{\Delta}{T_s}$ is quite large, after C_0 , the second sample itself will be not from the main lobe. However, if we choose $\frac{\Delta}{T_s}$ to not be too large yet not too small, we would have first few samples from main lobe, and we can observe that they will be of decreasing magnitude.

6.9.3. Sampling with the Pulse Train

We already know that the Fourier series expansion of $p(t)$ is $\sum_{k \in \mathbb{Z}} C_k e^{j \frac{2\pi}{T_s} kt}$, so multiplying it with $x(t)$, we'd have

$$x(t)p(t) = \sum_{k \in \mathbb{Z}} C_k x(t) e^{j \frac{2\pi}{T_s} kt}$$

Now let's take the Fourier transform of $x(t)p(t)$ which we have, assuming it exists. Note that, as Fourier transform is a linear transform, we can compute the Fourier transform of one term from the summation which we have, and add all such Fourier transforms. So we will compute Fourier transform of $x(t)e^{j \frac{2\pi}{T_s} kt}$.

$$\begin{aligned} \text{Fourier transform of } x(t)e^{j \frac{2\pi}{T_s} kt} &= \int_{-\infty}^{+\infty} x(t)e^{j \frac{2\pi}{T_s} kt} e^{-j\Omega t} dt \\ &= \int_{-\infty}^{+\infty} x(t)e^{-j(\Omega - \frac{2\pi}{T_s} k)t} dt \end{aligned}$$

6. chapter

Suppose we have Fourier transform of $x(t)$ as $X(\Omega)$, then the Fourier transform of $x(t)e^{j\frac{2\pi}{T_s}kt}$ will be $X\left(\Omega - \frac{2\pi}{T_s}k\right)$, from the above equation. i.e. it is the Fourier transform of $x(t)$ namely $X(\Omega)$ shifted by $\frac{2\pi}{T_s}k$ forward. So we have

$$\begin{aligned}\text{Fourier transform of } x(t)e^{j\frac{2\pi}{T_s}kt} &= X\left(\Omega - \frac{2\pi}{T_s}k\right) \\ \therefore \text{Fourier transform of } C_k x(t)e^{j\frac{2\pi}{T_s}kt} &= C_k X\left(\Omega - \frac{2\pi}{T_s}k\right) \\ \therefore \text{Fourier transform of } \sum_{k \in \mathbb{Z}} C_k x(t)e^{j\frac{2\pi}{T_s}kt} &= \sum_{k \in \mathbb{Z}} C_k X\left(\Omega - \frac{2\pi}{T_s}k\right)\end{aligned}$$

We can list down the procedure of finding the Fourier transform of $x(t)p(t)$ as follows:

1. Shift $X(\Omega)$ by $\frac{2\pi}{T_s k}$ forward, $\forall k \in \mathbb{Z}$.
2. Multiply this shifed version of $X(\Omega)$ by C_k .
3. Add up all these shifted versions.

6.9.4. Graphical Representation of the Fourier Transform of the Sampled Signal

We'll see how all of this looks graphically: Consider $x(t)$ to be bandlimited signal, as shown in the following image (magnitude of $X(\Omega)$ is shown in the image, and phase is considered to be zero for this signal everywhere):

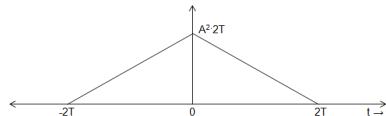
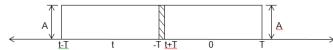


Figure 6.25.: $X(\Omega)$

Note that this Fourier representation corresponds to a real and even signal $x(t)$. We will first sample this signal 'disobeying' the Nyquist principle, i.e. by keeping $\frac{1}{T_s} \geq 2\frac{\Omega_m}{2\pi}$. The first few C_k values are seen in the following image:



Following the procedure given in the previous section, we obtain following spectrum for the sampled $x(t)$:

One feature we can clearly see in this graph is that carbon copies of the original spectrum are overlapping: which we had already seen when we had sampled a similar signal with impulse train instead of pulse train (ideal sampling). However,

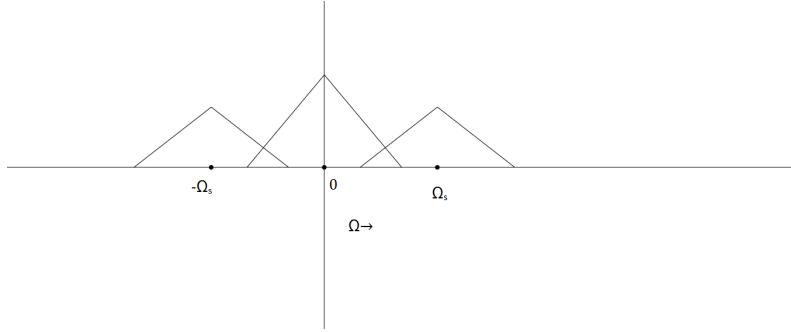


Figure 6.26.: Fourier transform when we disobey Nyquist principle: shows overlaps.

there's one more remarkable thing we can see here: As the first few C_k 's are decreasing in magnitude, we observe that the 'strength' of the carbon copies is reducing, at least for first few k 's. So the carbon copies still overlap, but their strength is lower, or, the magnitude of overlap is less now, the region in which such overlap is occurring is closer to the original spectrum, if compared with the Fourier transform of the same signal, sampled using impulse train.

Now we will see what happens if we sample this signal 'obeying;' the Nyquist principle, i.e. $\frac{1}{T_s} \gg 2\frac{\Omega_m}{2\pi}$. This time we obtain the spectrum of the sampled signal as following:

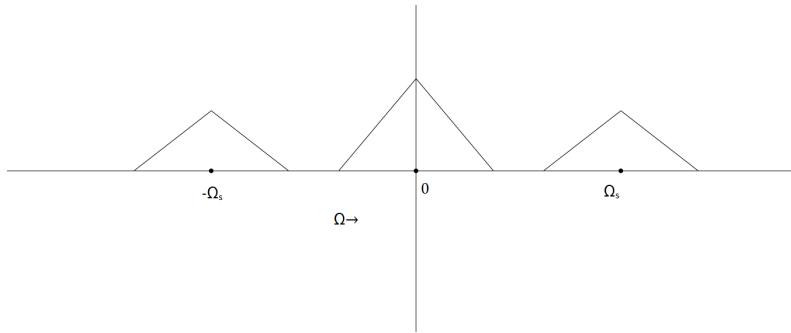


Figure 6.27.: Fourier transform when we obey Nyquist principle: no overlap is seen.

Here, no overlap is seen between the original spectrum and any of its carbon copies, similar to when we had sampled using impulse train. But we can still see the magnitude of the first few carbon copies decreasing.

When reconstructing the signal back in this case, still need our spectrum to lie in the 'flat' region of the spectrum of the reconstructor (a simple practical example of which, we have seen, is a simple RC - circuit). So practically Ω_s needs to be much higher than Ω_m . There is one more advantage of using pulse train instead of impulse train to sample a signal: for a practical reconstructor, it's spectrum doesn't become zero directly after the flat region- it gradually tends toward zero. So while reconstructing the signal, the higher-frequency carbon copies of the original spectrum aren't completely eliminated. Since we anyway have reducing magnitude for the carbon copies in the spectrum of the signal sampled using pulse train, we have lesser contribution from these high-frequency components than the reconstructed signal from the ideal sample of the signal.

6. chapter

We shall also give a name to these carbon copies: we will call them 'aliases' and this process 'aliasing' henceforth. Alias refers to the act of assuming a different identity. Here, original spectrum 'pretends' to be a spectrum around some Ω_s . If these aliases interfere with each otherm it only causes confusion, Nyquist theorem essentially tells us htat we don't want aliasing.

6.10. Lecture 10: Nyquist Sampling Theorem: Proof and Application

6.10.1. Introduction

Till now, we have talked about "ideal" sampling, where the sampling was done using impulses. We have also done sampling using *pulses*, instead of impulses. In this lecture, we will look at a "generalized version" of the sampling, which is called "amplitude modulation". We will also prove a more general form of the Nyquist-Shannon sampling theorem.

6.10.2. General form of the Nyquist-Shannon Sampling Theorem

We have looked at sampling in some detail in the previous lectures. We will now generalize this idea of sampling. Earlier, for "ideal" we multiplied the signal with a train of impulses placed at regular intervals. Now, what we are going to do is that we will multiply the signal with any "carrier signal" $p(t)$, and the only requirement that we are going to make of $p(t)$ is that it be periodic with period T_s . This signal, as a result of this periodicity, will have a complex Fourier series expansion:

$$p(t) = \sum_{k=-\infty}^{+\infty} c_k e^{\imath \frac{2\pi}{T_s} kt}$$

Let the signal to be "sampled" be $x(t)$, and I put the word "sampled" in quotes because what we're doing here can't really be called sampling. We need $x(t)$ to have a Fourier transform $X(\Omega)$. Now we "sample" this $x(t)$ using $p(t)$, (which is to say we multiply the two) to obtain $x(t)p(t)$. The Fourier Transform of this $x(t)p(t)$ is $X_s(\Omega)$ (say).

$$\begin{aligned} X_s(\Omega) &= \mathcal{F}(x(t)p(t)) \\ &= \mathcal{F}\left(x(t) \sum_{k=-\infty}^{+\infty} c_k e^{\imath \frac{2\pi}{T_s} kt}\right) \\ &= \sum_{k=-\infty}^{+\infty} c_k \mathcal{F}\left(x(t) e^{\imath \frac{2\pi}{T_s} kt}\right) \quad (k \text{ is independent of } t) \\ &= \sum_{k=-\infty}^{+\infty} c_k X\left(\Omega - \frac{2\pi}{T_s} k\right) \end{aligned}$$

As an example, let's see what happens with $p(t) = A_c \cos\left(\frac{2\pi}{T_s}t\right)$, which can also be written as

$$p(t) = \frac{A_c}{2} \left(e^{\imath \frac{2\pi}{T_s} t} + e^{-\imath \frac{2\pi}{T_s} t} \right)$$

Thus, we have

$$X_s(\Omega) = \frac{A_c}{2} \left\{ X\left(\Omega - \frac{2\pi}{T_s}\right) + X\left(\Omega + \frac{2\pi}{T_s}\right) \right\}$$

And this is *precisely* what is known as *amplitude modulation*. This $p(t) = A_c \cos\left(\frac{2\pi}{T_s}t\right)$ is called the *carrier wave*, and $x(t)$ is said to have been modulated with this carrier. This modulation has the effect of creating two copies of the signal and shifting each one of them forward and backward respectively, in the frequency domain. And if the carrier frequency is large enough, these two copies don't "overlap" (or "mix") with each other.

Therefore, in applying the Nyquist theorem, we have essentially carried out amplitude modulation (instead of sampling, as is usually done).

It's worthwhile to note that we need not have used a pure sinusoid to modulate the signal. In fact, as we have stated before, we could have used any periodic signal! This is important when we start designing and building systems (say electronic circuits), as it is difficult to build systems like multipliers and sinusoid generators, which are necessary for amplitude modulation with pure sinusoids.

Instead, a much simpler way to do modulation is to just use a train of pulses. It is easy to see that modulation using a pulse train (which is essentially just multiplication with said pulse train) is equivalent to just switching the signal on and off (and scaling it) at regular intervals! And building such a circuit would be much easier.

So we choose $p(t)$ to be a pulse train of period T_s , having pulse width Δ and height $\frac{1}{\Delta}$. Doing this, we obtain a spectrum as shown in Figure ???. As we can see, the first two carbon copies (the ones corresponding to c_1 and c_{-1}) are the ones we want, as they are the ones that appear in the usual amplitude modulation with a sinusoid.

We want to retain only the "carbon copies" corresponding to c_1 and c_{-1} and get rid of the c_0, c_2 and c_{-2}, c_3 and $c_{-3}\dots$ terms. To do that, we use a *band-pass filter* as shown in Figure ???.

A band pass filter, as the name suggests, only allows a certain *range* of frequencies to go through, and *blocks* all other frequencies. An ideal band-pass filter has a frequency response of G in the *passband* (the range of frequencies to be passed) and 0 in the *stop-band* (everywhere except the passband). The frequency response of a more *practical* band pass filter is shown in Figure ???. The frequencies at the boundary of the passband and stop band form the transition band. In this region, the impulse response undergoes a smooth transition. In a practical bandpass filter, you can't make the frequency response in the passband constant, you can't make the frequency response in the stop band exactly zero, and you can't make the transition band infinitesimally small.

Therefore, this filter *will* pass parts of the other copies of the original spectrum, but these will be *attenuated* (scaled down). Broadly speaking, this is how amplitude modulation is done, in practice.

6.11. Lecture 11: Sample and Hold

6.11.1. Introduction

In the last session we have seen that sampling is actually multiplying an input signal with a periodic waveform. Sampling is about representing the signal more efficiently than the raw form. In this discussion we discuss the problem of reconstruction of the signal if we put in place systems to interpolate the signal from the sampled signal.

6.11.2. Interpolation

Interpolate means to complete the waveform in between the samples. Interpolation is a method of constructing new data points within the range of a discrete set of known data points.

Graphical Explanation

Let us consider a function as shown in fig. 1 and sample it with a time period of T_s .

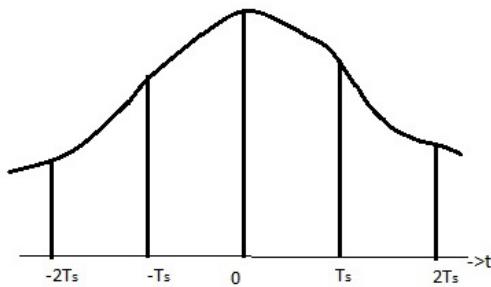


Figure 6.28.: A sampled function

There are various ways in which we can interpolate the given samples

Method 1: Sample and Hold

In figure 2, the red lines form the interpolated signal using the Sample and Hold method.

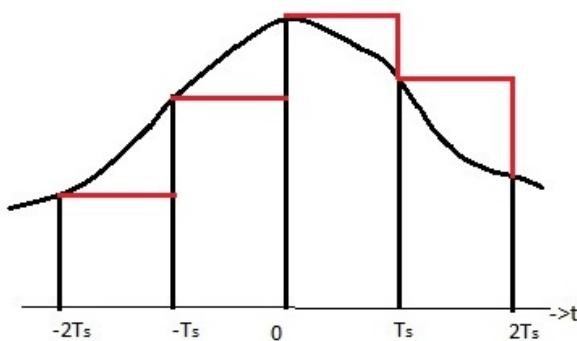


Figure 6.29.: Sample and Hold

Method 2: Linear interpolation

In this method the sampled points are connected by straight lines. Refer to figure 3 for linear interpolation. In Figure 3, the green lines form the linearly interpolated signal obtained by interpolating from the sampled signal.

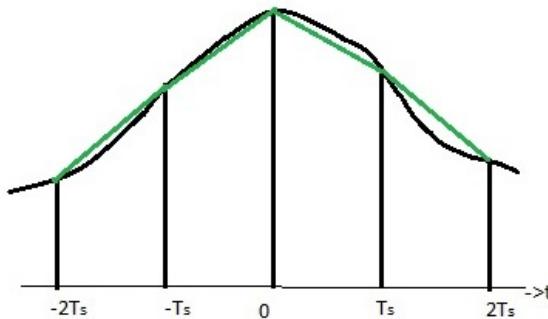


Figure 6.30.: Linear interpolation

Other kinds of interpolations are also possible. But in this discussion we shall focus on sample and hold.

Systematical explanation

Systematically the sample and hold can be understood as follows:

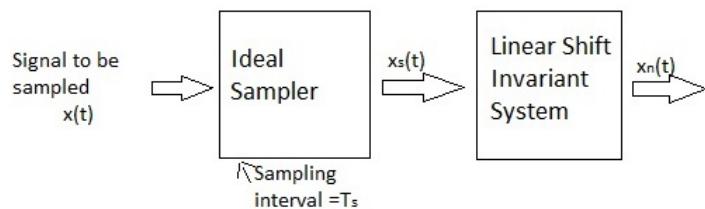


Figure 6.31.: Systematical representation of sample and hold

The impulse response of the linear shift invariant system in the above figure is a pulse as shown in fig 5.

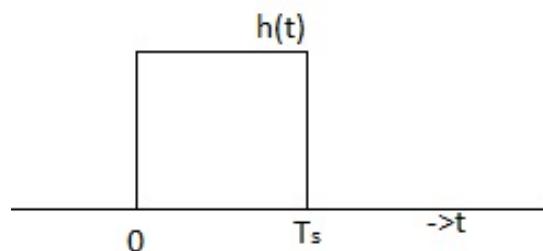


Figure 6.32.: Impulse response of linear shift invariant system.

6.11.3. Frequency domain analysis

The LSI system which follows the ideal sampler is called the Hold system. This is because, the effect of this system on a sampled signal is that of retaining the sampled value till the next sample is reached. Note that since the Hold system impulse response has a finite value in a finite interval, it is absolutely integrable and hence it is a stable system.

Now, the frequency response of the Hold system is:

$$\begin{aligned}
 H(\Omega) &= \int_{-\infty}^{\infty} h(t)e^{-jt\Omega} dt \\
 &= \int_0^{T_s} e^{-jt\Omega} dt \\
 &= \left. \frac{e^{-jt\Omega}}{-j\Omega} \right|_0^{T_s} \\
 &= \frac{1 - e^{-j\Omega T_s}}{j\Omega} \\
 &= \frac{e^{\frac{-j\Omega T_s}{2}} \left(e^{\frac{j\Omega T_s}{2}} - e^{\frac{-j\Omega T_s}{2}} \right)}{j\Omega} \\
 &= \frac{T_s}{2} e^{\frac{-j\Omega T_s}{2}} \frac{2j \sin \frac{\Omega T_s}{2}}{j \frac{\Omega T_s}{2}} \\
 &= T_s e^{\frac{-j\Omega T_s}{2}} \frac{2j \sin \frac{\Omega T_s}{2}}{2j \frac{\Omega T_s}{2}} \\
 &= T_s e^{\frac{-j\Omega T_s}{2}} \frac{\sin \Omega T_s / 2}{\Omega T_s / 2}
 \end{aligned}$$

Frequency domain analysis

Let $x(t)$ obey Nyquist principle. Hence band limited to $\frac{1}{2T_s}$. Consider such an $x(t)$ with Fourier Transform as shown in Figure 6.

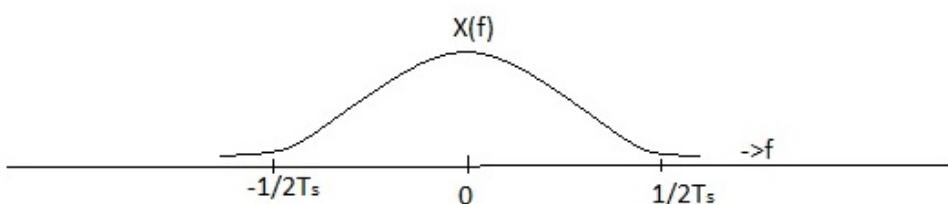
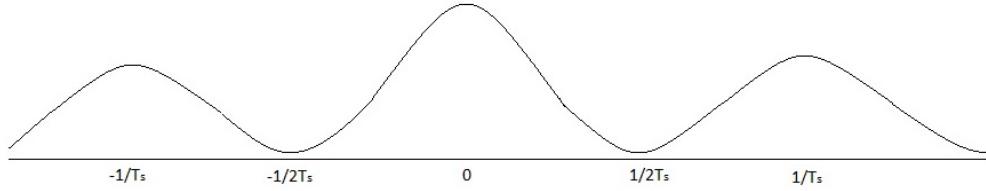


Figure 6.33.: Fourier transform of $x(t)$

Then, after passing the signal through the ideal sampler, the sampled signal obtained, $X_s(t)$, will have the spectrum $X_s(f)$ as shown in Figure 7.

Sketch of 'hold' frequency response.


 Figure 6.34.: Spectrum of the sampled signal: $X_S(f)$

$$H(f) = T_s e^{-j\pi f T_s} \frac{\sin \pi f T_s}{\pi f T_s}$$

Magnitude \Rightarrow

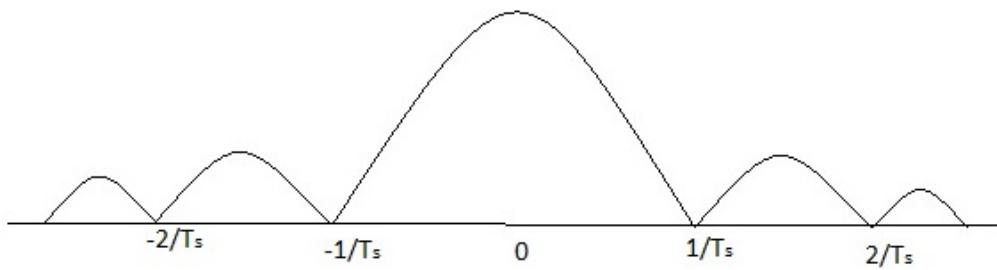
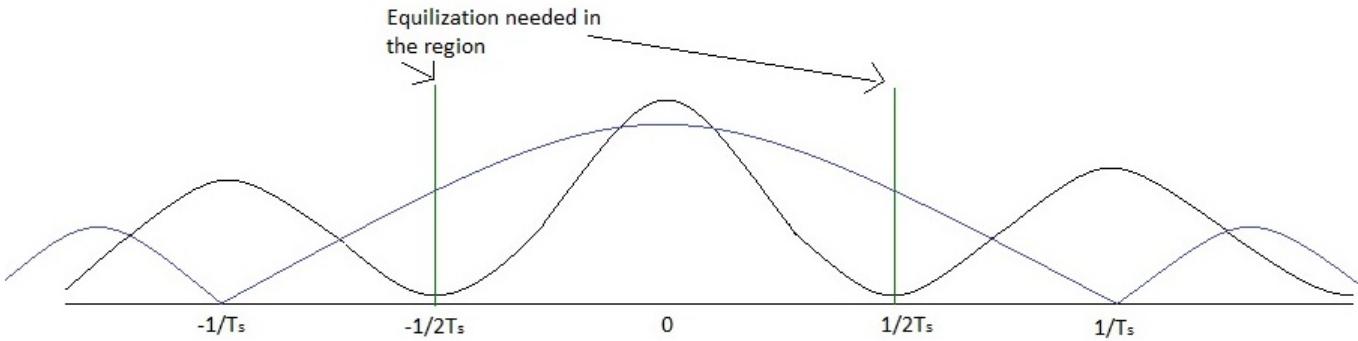


Figure 6.35.: Hold frequency response

If we superimpose both hold frequency response and $X_s(f)$,


 Figure 6.36.: Superimposing hold freq. response on $X_s(f)$

In process of sampling and holding, two problems are

1. Aliases pass through partially
2. Original gets distorted

Hence, equalization is needed in the original spectral band of $-\frac{1}{2T_s} \rightarrow \frac{1}{2T_s}$

Equalization is the reciprocal of the hold frequency response given by

$$\frac{1}{T_s e^{-j\pi f T_s} \frac{\sin \pi f T_s}{\pi f T_s}} \text{ for } -\frac{1}{2T_s} < f < \frac{1}{2T_s}$$

6. chapter

Sketch of equalizer frequency response.

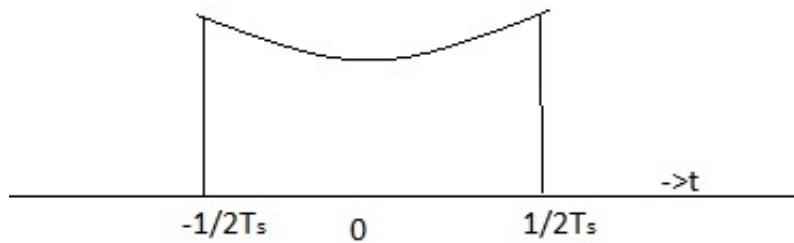


Figure 6.37.: Equalizer freq. response

Hence to reconstruct a signal after sample and hold, two things are required:

1. To remove the aliases
2. To equalize

The ideal reconstructor for sample and hold has to

1. Equalize in signal band
2. Stopband outside signal band

In the next lecture, we will discuss about linear interpolation.

6.12. Lecture 12: Sampling followed by Linear Interpolation

6.12.1. Introduction

In the previous section, we have seen the working of a Sample and Hold circuit. We can now say that the Sample and hold circuit is nothing but an Ideal Sampling System followed by a Linear Shift Invariant system.

We have also seen the drawbacks of the sample and hold circuit which include:

1. Distortion in the original spectrum.
2. Aliases also pass through the circuit partially.

We will now look at the procedure that will be followed by the Sample and Hold circuit.

6.12.2. Linear Interpolation

Linear Interpolation can be broadly defined as the joining of two known points by a straight line.

Thus, if we know two given points, the linear interpolant will be the straight line joining these two lines.

Above figure indicates a rough sketch of Linear Interpolation of a given set of points.

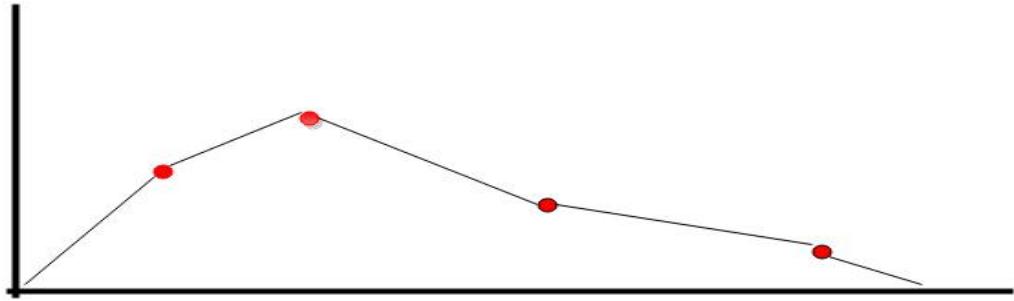


Figure 6.38.: Linear Interpolation of given points.

6.12.3. Signal Representation of Linear Interpolation

We now look at the effect of the Linear Interpolation on a given signal $x(t)$.

Signal $x(t)$ is passed through an ideal sampler, with sampling interval T_s . It is then passed through a Linear Shift Invariant system.

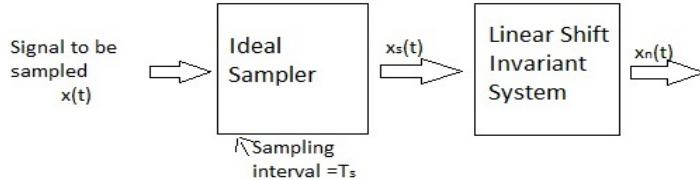


Figure 6.39.: Passing $x(t)$ through the Sample and Hold Circuit.

The impulse response of the Linear Shift Invariance will be triangular in nature as shown below. We sketch the original signal $x(t)$, the ideal sampling output $x_s(t)$ and the scaled and shifted impulse response of the Linear Shift Invariant System given by $x(nT_s).h(t - nT_s)$

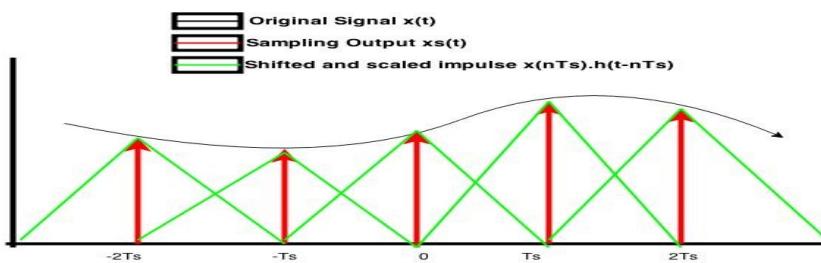


Figure 6.40.: Signal Representation along with the Sampling output and the Impulse response of LSI system

The shifted and scaled impulse responses are now added linearly. As seen before, linear interpolation is the joining of two points with a straight line.

We now add the *green* straight lines between the various time intervals. Additions of two straight lines will also be a straight line.

The resulting addition of these straight lines will be the Linear Interpolated Signal and can be seen in *blue* in the figure below.

6.12. Lecture 12: Sampling followed by Linear Interpolation

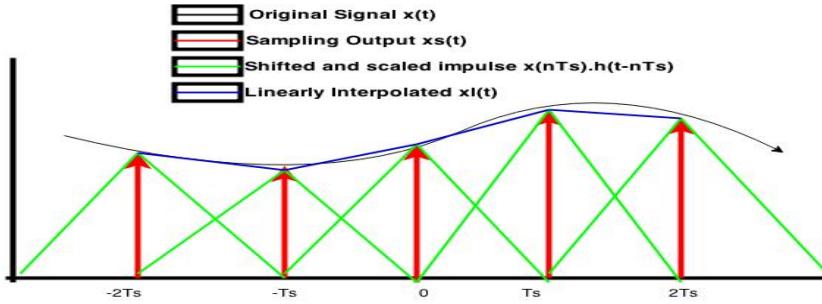


Figure 6.41.: Signal Representation along with the Sampling output and the Impulse response of LSI system

Idea of Linear Interpolation for a signal can be stated as the passing of an ideally sampled output through a Linear Shift Invariant system.

6.12.4. Significance in frequency domain

Convolution of a signal with itself can be seen as the multiplication of it's Fourier Transform with itself.

The sample signal $h_0(t)$ can be shifted to the time interval $-Ts/2$ to $Ts/2$. This gives us the Zero order hold.

Convolution $h_0(t)$ with itself gives us the first order hold. It can be seen in the given diagram.

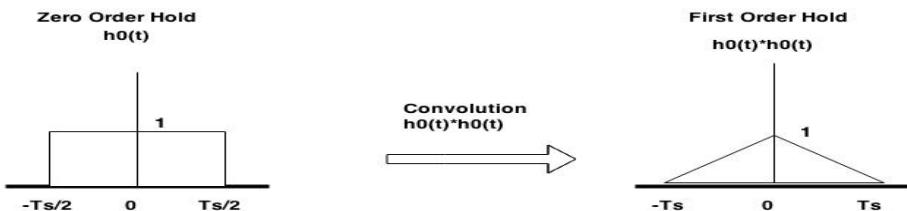


Figure 6.42.: Convoluting the Zero order hold with itself

Mathematical representation

Frequency Response of first order hold:

$$Ts \cdot \exp(-jt\omega) \cdot \frac{\sin(\frac{Ts}{2}\omega)}{\frac{Ts}{2}\omega} \cdot \exp(\frac{Ts}{2}\omega)$$

Squaring the above term, we get:

$$Ts^2 \cdot \left(\frac{\sin(\frac{Ts}{2}\omega)}{\frac{Ts}{2}\omega} \right)^2$$

The implications will be as follows:

1. $\text{sinc}(fTs)$ has a maximum value of 1.
2. We know that the function $\text{sinc}(fTs) < 1$. Therefore, squaring the function will result in further reduction of magnitude.
3. The function will thus fall at a faster rate.

We can thus conclude from above that the given operation has the following effects on the signal band:

Advantages We can thus see that the side lobes are now lower. The other aliases that were formed are suppressed as compared to the Zero order hold.

Disadvantage Along with the side lobes, the major lobe too will now be suppressed to some extent. The region from $-Ts/2$ to $Ts/2$ suffers a steeper drop.

6.13. Lecture 13: Applications of Discrete Time Processing

6.13.1. Introduction

In the previous lectures in Module 3, we have studied the concepts of sampling of continuous independent variable signals and reconstruction from the sampled signals. In some ways, we thus converted a continuous signal into a discrete one in the process of sampling. Now we wish to apply these concepts learnt to practical scenarios. We motivate this application below.

6.13.2. Why Digital Systems over Continuous Systems?

Digital vs Discrete Signals

We first bring out the difference between a digital system and a discrete system. A discrete signal means that the independent variable takes on discrete values rather than continuous values. This we already know. Thus sampling a continuous signal leads to a discrete signal. The result of sampling, a discrete signal, however, takes on arbitrary real values. Implementing this in practice is not possible as we can only represent fixed precision numbers in all our processing devices. Thus, we need to discretise the values taken by the sampled signal i.e. we need to discretise the dependent variable too. The result is then called a digital signal. We give an example below.

Discrete vs Continuous Signals

We have already talked about how discrete signals (and digital signals more so) are convenient for processing and also that they are robust against noise. Somewhat similar is the case with digital and continuous systems. We list below some advantages of digital systems over continuous systems.

1. Flexibility

Digital systems are generally implemented on a digital signal processor (DSP) which, in layman's terms, basically are just computers. The jargon just makes it sound frightening. As is the case with a computer, even a DSP can be made to do a lot of diverse things without much effort. Thus, we just need to change the program and we get a new system implemented without much effort. That however, is not the case with an analog system as it is generally highly specialized and hence restrictive.

2. Easy and Low Maintenance

Generally, digital circuit components have a longer life as compared to analog components. Also, a digital system, is by construction more modular as compared to an analog one leading to easy replacement of parts when necessary.

3. Generality

Designing an analog system for any given frequency response is often challenging and even practically impossible. This however is not the case with digital systems where we can get as close to the desired response as needed with more processing and possibly more memory.

6.13.3. A Digital System

From the above discussion, it is clear that digital systems are preferred over continuous systems practically. However, the practical world is an analog one and hence the signals there are continuous signals. To use a digital system to process a real world signal, we need to convert the signal into a digital signal, then process it using a DSP and then reconverit it back to an analog signal for it to be of any use to us. This architecture is shown in figure 1.

As is seen in the figure, we need to sample the input signal (discretise the independent variable) and later quantise it (discretise the dependent variable). Sampling necessitates that the input signal be bandlimited. This is in general not an issue because most of the signals in real life are practically bandlimited given a sufficient frequency range.

6.13.4. A Speech Processing System

We now take a concrete example of a speech signal and discuss the above structure of a digital system in that context. As we know, a speech signal is generally bandlimited to $4kHz$. In that case, the Nyquist Sampling Theorem tells us that the signal be sampled at a sampling frequency of more than $8kHz$. We also, however, know that a margin is better for ease of reconstruction. Thus we decide the sample the speech signal at a frequency of $10kHz$ i.e. we record 10,000 samples per second.

6. chapter

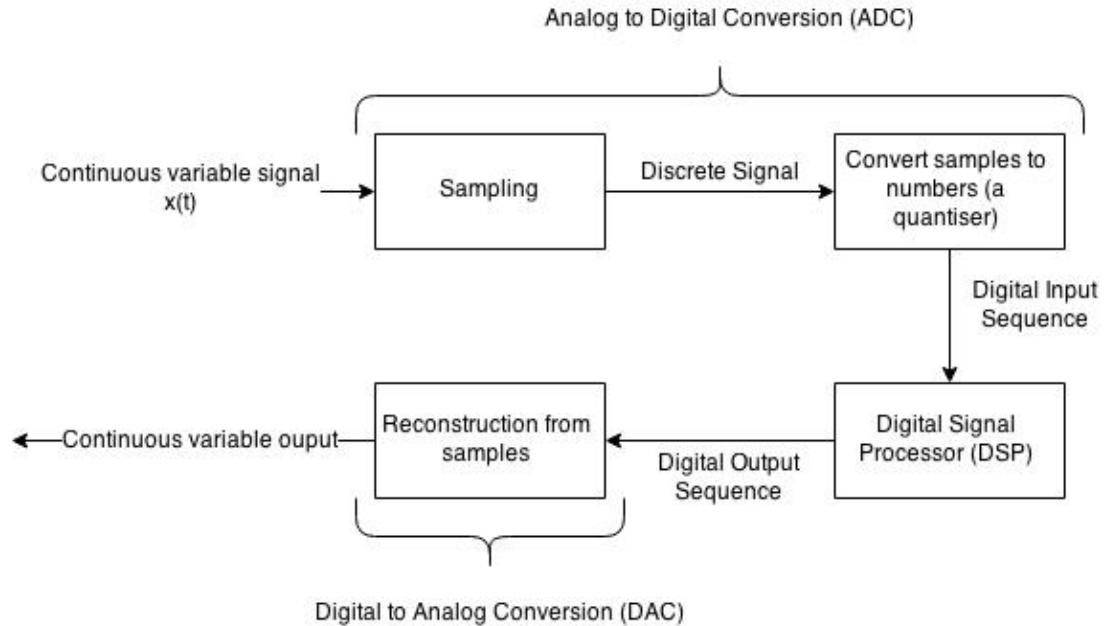


Figure 6.43.: Processing a Continuous Signal Using a DSP

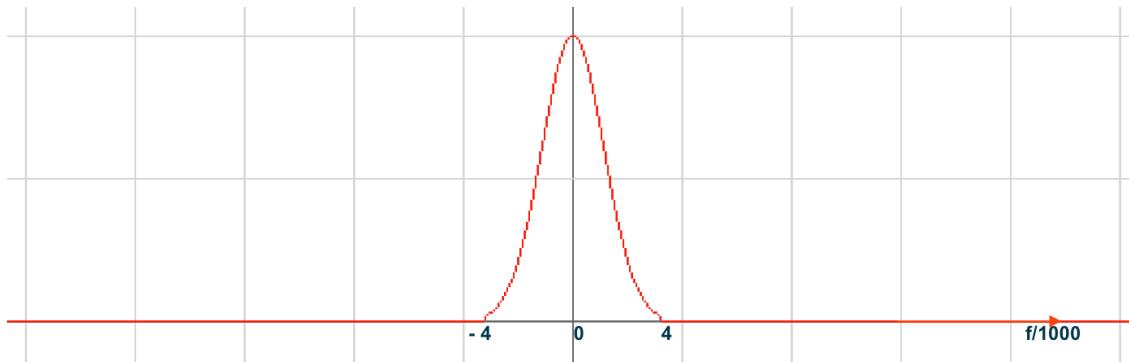


Figure 6.44.: Spectrum of a Speech Signal

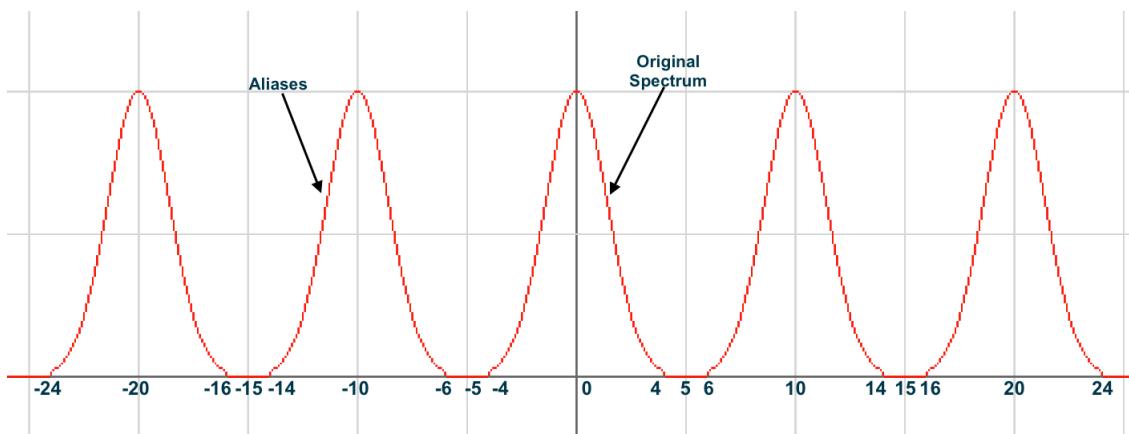


Figure 6.45.: Spectrum of the Sampled Speech Signal

Let the frequency spectrum of the speech signal be as shown in figure 2(a) (Note

that this is not the spectrum of a real speech signal - which is a lot more complicated - but an illustrative plot) On sampling this signal at 10kHz , we get the spectrum shown in figure 2(b).

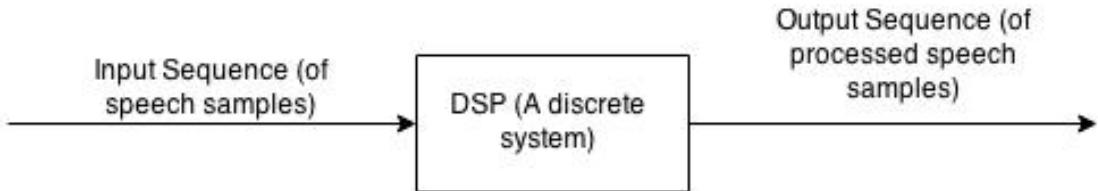


Figure 6.46.: A Digital System

We then pass these digital samples (which now are just bits) into the digital signal processor to process the signal as desired. A schematic of this discrete system is shown in figure 4. A discrete system takes as input the input samples and concurrently gives a sequence at the output. After this processing is done, we reconstruct an output signal from the samples from the DSP using a DAC as shown in figure 1.

6.13.5. An Overview of a Successive Approximation ADC

Here we present an overview of a successive approximation analog to digital converter. A block schematic is given in figure 5. As the name suggests, we start with an approximate binary value of the input voltage and successively (iteratively) improve this approximation. We convert the current approximation of the input (a binary value) into an analog signal with the help of a voltage combiner (this is just a DAC which we look at in the next section). The output of this combiner and the input voltage are then passed to a comparator. Depending on the result of the comparator, we increase or decrease the value of the current approximation (i.e. we improve the approximation in this step). This continues till we can no longer improve the approximation because of the limited precision of the finite number of binary bits available to us.

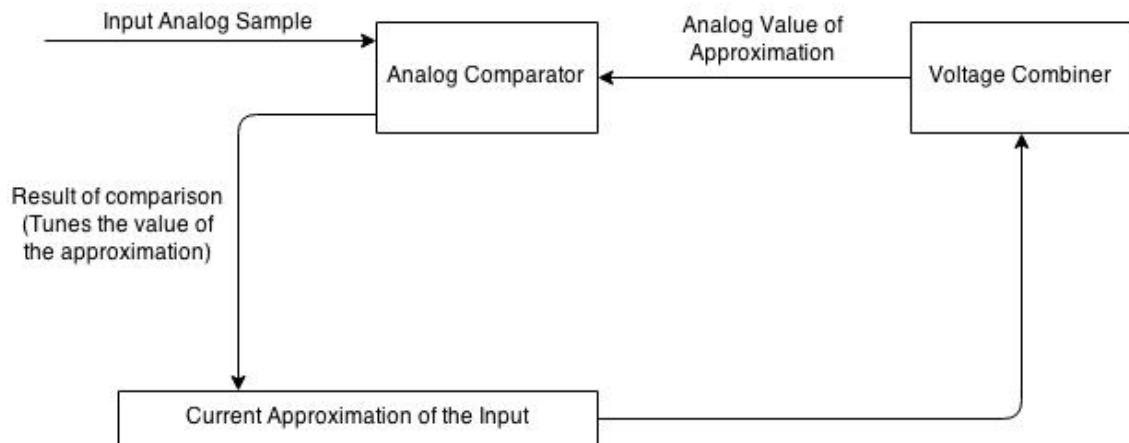


Figure 6.47.: Successive Approximation ADC

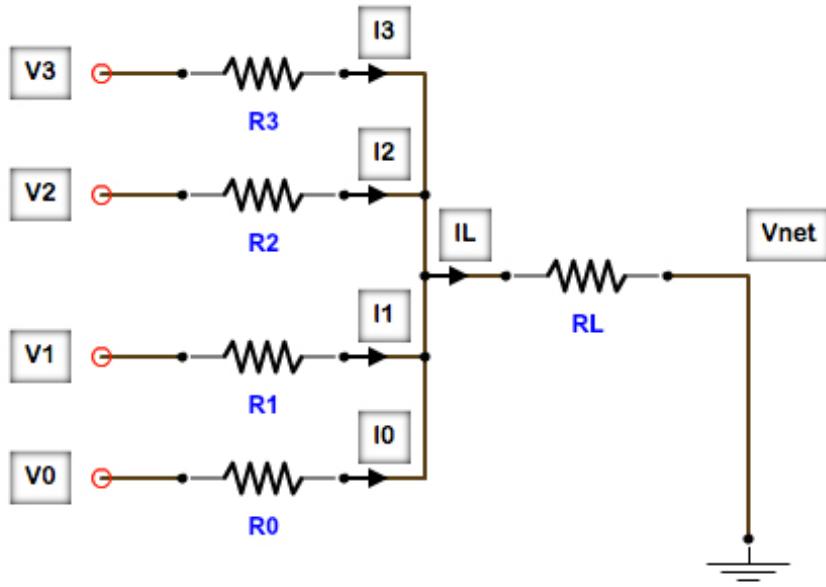


Figure 6.48.: A Possible DAC

6.13.6. An Overview of a DAC

Now, we describe a Digital to Analog Converter (DAC). In general, a DAC takes a n -bit binary number and converts it into a voltage proportional to the value of the input binary number. We consider the case of a 4-bit DAC as shown in figure 6. V_0 , V_1 , V_2 and V_3 are the 4 bits (We assume that the bit 1 corresponds to a voltage of 5V and the bit 0 corresponds to a voltage of 0V.). R_0 , R_1 , R_2 , R_3 and R_L are resistors such that $R_1 = 2R_0$, $R_2 = 4R_0$ and $R_3 = 8R_0$. As shown, I_0 , I_1 , I_2 , I_3 and I_L are currents through the resistors R_0 , R_1 , R_2 , R_3 and R_L respectively.

We have, by Ohm's Law,

$$I_0 = \frac{V_{net} - V_0}{R_0}$$

$$I_1 = \frac{V_{net} - V_1}{R_1}$$

$$I_2 = \frac{V_{net} - V_2}{R_2}$$

$$I_3 = \frac{V_{net} - V_3}{R_3}$$

$$I_L = \frac{V_{net}}{R_L}$$

In that case, by Kirchoff's Current Law (KCL), we have,

$$I_L = I_0 + I_1 + I_2 + I_3$$

Thus,

$$\begin{aligned}\frac{V_{net}}{R_L} &= \frac{V_{net} - V_0}{R_0} + \frac{V_{net} - V_1}{R_1} + \frac{V_{net} - V_2}{R_2} + \frac{V_{net} - V_3}{R_3} \\ &= \frac{V_{net} - V_0}{R_0} + \frac{V_{net} - V_1}{2R_0} + \frac{V_{net} - V_2}{4R_0} + \frac{V_{net} - V_3}{8R_0}\end{aligned}$$

Rearranging,

$$V_{net} \left(\frac{1}{R_L} + \frac{1}{8R_0} + \frac{1}{4R_0} + \frac{1}{2R_0} + \frac{1}{R_0} \right) = \frac{V_3}{8R_0} + \frac{V_2}{4R_0} + \frac{V_1}{2R_0} + \frac{V_0}{R_0}$$

If,

$$G_{L1} = \left(\frac{1}{R_L} + \frac{1}{8R_0} + \frac{1}{4R_0} + \frac{1}{2R_0} + \frac{1}{R_0} \right)$$

and

$$G_0 = \frac{1}{R_0}$$

Here G stands for conductance, the inverse of resistance. Then we have,

$$V_{net} = \frac{G_0}{G_{L1}} \frac{1}{8} V_3 + \frac{G_0}{G_{L1}} \frac{1}{4} V_2 + \frac{G_0}{G_{L1}} \frac{1}{2} V_1 + \frac{G_0}{G_{L1}} V_0$$

Thus, if V_0 corresponded to the Most Significant Bit (MSB) of the input binary number and V_3 corresponded to the Least Significant Bit (LSB), then V_{net} is proportional to the value of the input binary number as is seen. The proportionality constant does not matter a lot because we can always amplify a given voltage if necessary.

6.13.7. Frequency Normalisation

Before we start discussing concrete discrete (or digital) systems, we need get a few boring things out of the way. Notation and convention is one of them. We discuss here something called frequency normalisation. This basically means that we look at quantities all relative to the sampling frequency (or equivalently the sampling interval). Thus, we consider one sampling interval to be a unit length of time. The sampling frequency, which is equal to the reciprocal of the sampling interval (which is now equal to one unit of time) also is equal to unity in these normalised units. Thus, in order to avoid aliasing, according to Nyquist's theorem, the maximum frequency component in the original signal should be less than 0.5. Similarly, we can also normalise the angular frequency. This will be equal to 2π times the normalised cycles-per-second frequency. We will use these normalised frequencies in the future lectures to conveniently study more about discrete systems and related concepts.

An example

Consider a system that takes an input signal as an input and outputs only the positive part of it i.e.,

$$y(t) = \begin{cases} 0 & x(t) \leq 0 \\ x(t) & x(t) > 0 \end{cases}$$

In order to implement this system in the analog domain, we would use a diode. The current-voltage characteristics of a diode (for low values of the voltage) are shown in the figure 8 below.

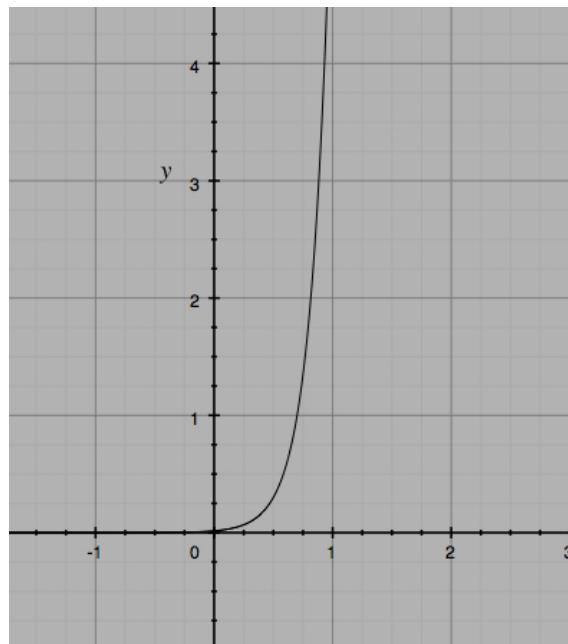


Figure 6.49.: A Possible DAC

The circuit one would use for this purpose is thus shown in figure 7.

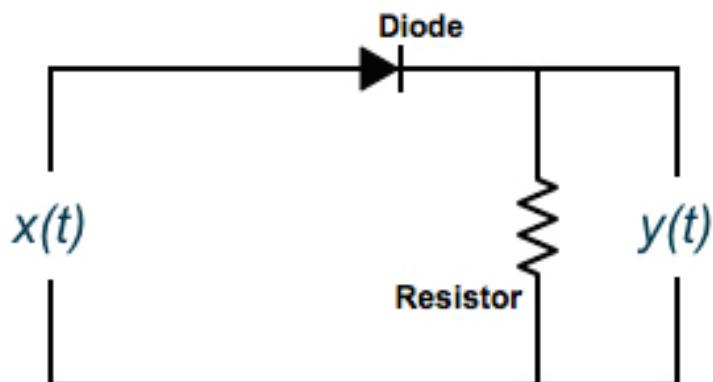


Figure 6.50.: A Possible DAC

As can be seen from the circuit, the output voltage will be equal to (according to the Shockley diode equation, about which you do not have to know anything :-))

$$y(t) = nV_T \log\left(\frac{x(t)}{I_s R} + 1\right)$$

Here, n , V_T and I_s are some constants characterizing the diode and R is the resistor in the circuit. The point however, is that the output is not exactly the same as the input when the input is positive and is not exactly zero when the input is negative though it will be close to what is desired. In order to improve the performance, we might want to use better diodes which will lead to a rise in the cost of the system.

In the digital domain, however, designing a system to achieve what is desired is both easy and will provide good results if the conversion from the analog to digital domain and back is done properly (which can be done by increasing the sampling rate and decreasing the quantisation interval). Thus, one can see that designing (and improving) a system in the digital domain is much easier than in the analog domain. Also, digital systems provide better results as is seen because unlike analog systems, they are not affected by non idealities of components. (in this case, the diode is non ideal.)

6.14. Lecture 14: Discrete Time Fourier Transform

6.14.1. Introduction

We have so far talked about building a speech system, which would operate/process the signal in discrete time. For that, we first sample our periodic hence bandlimited signal, obeying the Nyquist principle. The resultant spectrum looks like this: (Specific signal chosen could have a different bandlimited spectrum than this, we've shown it to be linear just as an example)

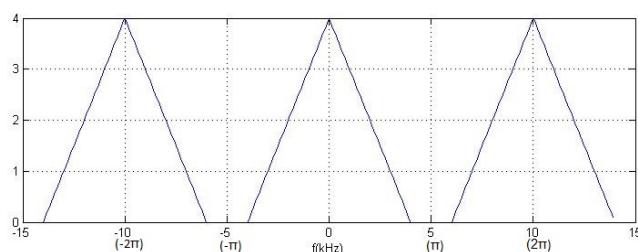


Figure 6.51.

Observe that the original spectrum exists here, along with its carbon copies or aliases. Now we change our x-axis to show normalized angular frequency, ω instead of cycles per second frequency as shown. In our example, the signal $x(t)$ is bandlimited by 4 kHz, and our sampling frequency is 10 kHz. So in the normalized angular frequency, 5 kHz will correspond to ' π ', 4 kHz to ' 0.8π ' and so on, as shown.

6.14.2. Discrete Time Fourier Transform

Since we are processing the signal $x(t)$ in discrete time, our system along with signals can be represented as following:



Figure 6.52.

We'll choose our system to be the one described by $y[n] = \frac{x[n]+x[n-1]}{2}$. But for doing this, first we need to sample $x(t)$ to obtain $x[n]$. We have also decided to normalize our units, i.e choosing sampling interval as our unit of time, and sampling frequency as $f_s = 1$ (for cycles/samples per second frequency) or $\Omega_s = 2\pi$ (for angular frequency). In the chosen example, we have signal $x(t)$, which is a speech signal, sampled at 10 kHz i.e. taking 10,000 samples in each second. So our samples can be given as $x(nT_s)$, where $n \in \mathbb{Z}$ and $T_s = 0.1\text{ms}$.

So formally, the signal after sampling is given as

$$\begin{aligned} \text{sampled signal} &= x(t) \sum_{n=-\infty}^{+\infty} \delta(t - nT_s) \\ &= \sum_{n=-\infty}^{+\infty} x(nT_s) \delta(t - nT_s) \end{aligned}$$

This is the signal which we'll give as input to the system S . Now the Fourier transform of our sampled signal will be

$$\text{FT of sampled signal} = \int_{-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} x(nT_s) \delta(t - nT_s) e^{-j\Omega t} dt$$

(If this integral converges, and we'd expect it to, for a speech signal, we'll exchange summation and integration) Hence

$$\begin{aligned} \text{FT of } x[n] &= \int_{-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} x(nT_s) \delta(t - nT_s) e^{-j\Omega t} dt \\ &= \sum_{n=-\infty}^{+\infty} x(nT_s) \int_{-\infty}^{+\infty} \delta(t - nT_s) e^{-j\Omega t} dt \\ &= \sum_{n=-\infty}^{+\infty} x(nT_s) e^{-j\Omega nT_s} \end{aligned}$$

This, as we already know, represents the spectrum shown in this lecture. In terms

of our normalization, we have,

$$\begin{aligned}
 x(nT_s) &= x[n] \\
 \text{also, } e^{-j\Omega nT_s} &= e^{-j\left(\frac{\Omega}{T_s}\right)n} \\
 \text{But } \frac{\Omega}{\left(\frac{1}{T_s}\right)} &= \frac{2\pi \times \text{Actual cps frequency}}{\text{Sampling Frequency}} \\
 &= \text{Normalized Angular Frequency, } \omega \\
 \therefore e^{-j\Omega nT_s} &= e^{-j\omega n}
 \end{aligned}$$

$$\begin{aligned}
 \text{FT of sampled signal} &= \sum_{n=-\infty}^{+\infty} x(nT_s)e^{-j\omega n} \\
 &= \sum_{n=-\infty}^{+\infty} x[n]e^{-j\omega n}
 \end{aligned}$$

This is essentially the Discrete Time Fourier Transform (DTFT) of $x(t)$ or of sequence $x[n]$. It is represented by $X(\omega)$. We can easily observe that it is essentially the component of $x[n]$ along $e^{j\omega n}$ /inner product of $x[n]$ and $e^{j\omega n}$. It tells us how much of $e^{j\omega n}$ is in the signal for each n .

As ω is the normalized angular frequency, spectrum of our signal would be upto $\omega = \pi$ strictly speaking. However, we shouldn't infact go upto $\omega = \pi$ and leave some margin as a buffer. For example, in the signal we have considered, $\omega = \pi$ corresponds to 5 kHz, but our signal only goes up to 4 kHz, or $\omega = 0.8\pi$.

6.14.3. Processing the Discretized Signal



Figure 6.53.

We had our system description as $y[n] = \frac{x[n]+x[n-1]}{2}$. Let us analyse what happens to a sinusoidal signal input. We are choosing sinusoid as we already know how to decompose a signal (if at all it is possible) in terms of sinusoids, and we have all the tools studied during our foray into Fourier domain analysis to help us. Speech signal, in particular, is usually bandlimited by 4 kHz, where lower frequencies are usually of the male voices (0-2 kHz) and female voices generally occupy higher frequencies (2-4 kHz). In our speech signal, we've considered a mixture of both male and female voices, as a general case. And we have sampled this signal at 10 kHz.

For analysis, we will use the rotating complex number representation of sinusoidal signal, $e^{j\omega n}$. So we have input to our system $x[n] = e^{j\omega n}$. And as $e^{j\omega n}$ is an eigen

6. chapter

sequence, we know that a multiple of itself is going to come out as output, i.e. output will be of the form ‘constant $\times e^{j\omega n}$ ’. Using the system description we have,

$$\begin{aligned} y[n] &= \frac{x[n] + x[n-1]}{2} \\ &= \frac{e^{j\omega n} + e^{j\omega(n-1)}}{2} \\ &= \frac{1}{2}e^{j\omega n}(1 + e^{-j\omega}) \end{aligned}$$

Hence the constant multiplying $e^{j\omega n}$ is $\frac{1}{2}(1 + e^{-j\omega})$. This constant can be simplified as:

$$\begin{aligned} \frac{1}{2}(1 + e^{-j\omega}) &= \frac{1}{2}e^{-j\omega/2}(e^{j\omega/2} + e^{-j\omega/2}) \\ &= \frac{1}{2}e^{-j\omega/2}(2 \cos(\omega/2)) \\ &= e^{-j\omega/2} \cos(\omega/2) \end{aligned}$$

In this form, we can easily identify the magnitude and phase factor, by which $e^{j\omega n}$ is multiplied. As magnitude of $e^{-j\omega/2}$ is one, it contributes only to the phase change, and as $\cos(\omega/2)$ corresponds to the magnitude change, and not to phase change, as in our range of $-\pi$ to π , $\cos(\omega/2)$ is a non-negative real number (hence has zero phase). This magnitude change can be plotted graphically in terms of ω as

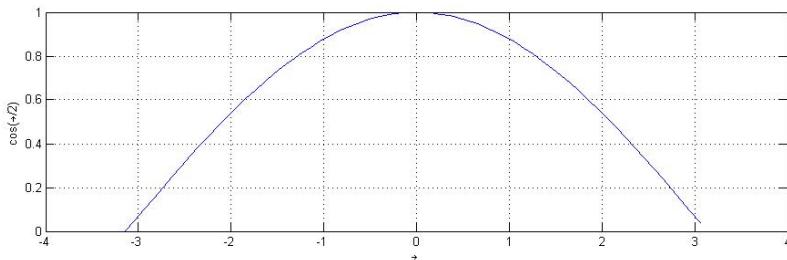


Figure 6.54.

Remembering that 0-2 kHz represents predominantly the male voices, and 2-4 kHz the female voices, we see that this constant has higher magnitude for male voices, in general, than for the female voices. So this system enhances the male voices and suppresses the female voices.

The student is encouraged to try and find a simple system doing just the reverse thing, suppressing the male voices and enhancing the female voices. As a start, try considering $y[n] = \frac{x[n]-x[n-1]}{2}$, and giving it a rotating complex number as an input, $e^{j\omega n}$, to observe that the output is a multiple of this input, given by $\frac{e^{j\omega n}+e^{j\omega(n-1)}}{2} = \frac{1}{2}e^{j\omega n}(1 + e^{-j\omega})$. Analyse the constant to deduce that this system does the required job.

6.14.4. Pondering About DTFT

We observed that the DTFT of a signal is given by $X(\omega) = \sum_{n=-\infty}^{\infty} x[n]e^{-j\omega n}$. We have confined ω within a region of length π . Let us try shifting this by 2π :

$$\begin{aligned} X(\omega + 2\pi) &= \sum_{n=-\infty}^{\infty} x[n]e^{-j(\omega+2\pi)n} \\ &= \sum_{n=-\infty}^{\infty} x[n]e^{-j\omega n}e^{-j2\pi n} \\ &= X(\omega), \text{ as } e^{-j2\pi n} = 1 \forall n \end{aligned}$$

Hence, shift of 2π make no difference, and this is what we'd expected, as DTFT is the spectrum of an ideally sampled signal, although written on normalized angular frequency axis. The spectrum of a discrete time sampled signal is always periodic, with a period of sampling frequency, and as we have normalized sampling frequency to 2π , it is now periodic with 2π . Also, any operations which we perform (as they are also in discrete time, with same time interval) are also periodic with the same interval of 2π , and would act not only on the original spectrum, but also the carbon copies.

Also, we have done this analysis for signals sampled with ideal impulse-train sampling. A similar deed may be performed on signals sampled with pulse train sampling also, the case in which the carbon copies would have reduced magnitudes, but we'll not go there; however interested students are encouraged to try and think over it.

6.15. Lecture 15: Discrete time signal processing system

6.15.1. Introduction

In this lecture, we will see how we can take all the processing that we wish to do on continuous-time signals into the discrete-time domain. This can be easily done on a computer.

6.15.2. General Discrete-time processing

We will be dealing with continuous signals which have a Fourier transform and systems which have a frequency response. From our previous discussions on sampling and reconstruction, we know that we need continuous signals to be *band-limited* if we wish to perfectly reconstruct them after sampling. Hence we will limit ourselves to those continuous-time signals whose Fourier transform is zero beyond $|\Omega| \geq \Omega_m$.

Recall that every type of sampling ultimately means that there is a repetition of the spectrum of the original signal at every integral multiple of the sampling frequency. The sampling frequency has to be *greater* than twice the maximum frequency at which the signal is non-zero (in order to have no aliasing or overlap between any two carbon copies of the original spectrum). That is, $f_s > 2\frac{\Omega_m}{2\pi}$. This is the *analog-to-digital conversion*. From Nyquist's principle, it is then possible to perfectly reconstruct the signal from its samples. This is called *digital-to-analog conversion*.

But there is no purpose to simply sampling and reconstructing. The beauty in this scheme is when we process the samples i.e when we do something between the analog-to-digital conversion and digital-to-analog conversion. This is called discrete-time signal processing (DSP). As we will be dealing with band-limited signals, it is enough to restrict the discrete-time systems also to this band.

Let the input band-limited signal have a Fourier transform $X(\Omega)$ and the output signal have a Fourier transform $Y(\Omega)$. If the frequency response of the LSI system is $H(\Omega)$, then we have $Y(\Omega) = X(\Omega)H(\Omega)$. Let us look at a consequence of this in Fig. 6.55. Because of the previous relation, $Y(\Omega)$ is zero whenever $X(\Omega)$ or

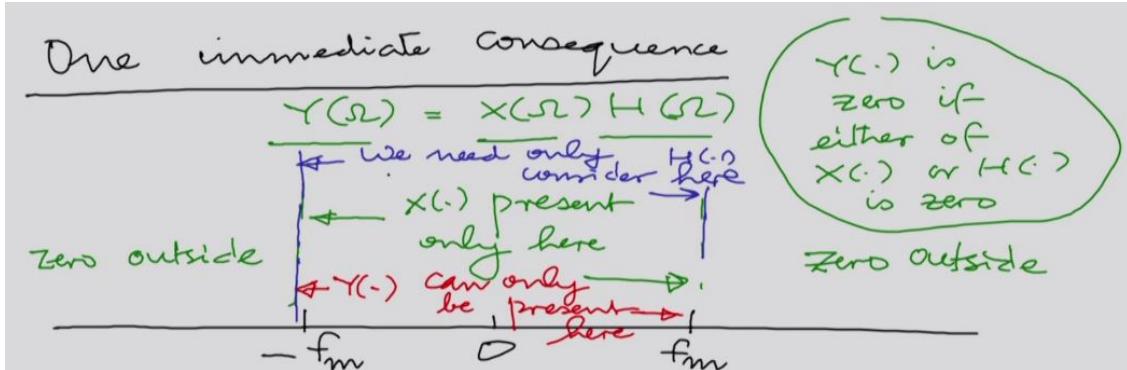


Figure 6.55.: Region of relevance for $Y(\Omega)$ and $H(\Omega)$

$H(\Omega)$ is zero. Hence because of the band-limited nature of $X(\Omega)$, we can ascertain that $Y(\Omega)$ is also band-limited between $-\Omega_m$ and Ω_m . As a consequence, we need consider $H(\Omega)$ in the same band (shown in blue in Fig.6.55).

6.15.3. Going from continuous-time to discrete-time signal processing

A discrete-time signal processing system does the following:

1. It takes a continuous-time band-limited signal $x(t)$, band-limited as $-\Omega_m \leq \Omega \leq \Omega_m$
2. It samples it at a rate f_s , which as per the Nyquist principle has to be greater than $2\frac{\Omega_m}{2\pi}$ which yields the input sequence $x[nT_s]$. For convenience, we denote this as $x[n]$ by calling T_s as the unit time. This is called as *normalisation*.
3. This input sequence $x[n]$ is fed into a linear shift-invariant system with the impulse response $h[n]$, to yield the output sequence $y[n]$.

Our aim is to process a continuous-time signal $x(t)$ by operating a system (frequency response = $H(\Omega)$) on it. We want to devise a discrete-time system which can achieve the same effect when we operate it on $x[n]$ and reconstruct the resulting $y[n]$. Let us look at this in more detail.

We have a continuous-time band-limited signal $x(t)$. We operate (process) a continuous-time LSI system with frequency response $H(\Omega)$, to obtain $y(t)$. The output signal has to be band-limited. We are quite safe if we restrict $H(\Omega)$ to $-\Omega_m \leq \Omega \leq \Omega_m$. This gives us the discrete system that we are looking for. In

other words, we can treat the restricted version of the frequency response of $H(\Omega)$ as the frequency response of the discrete-time system.

Important!- The discrete-time Fourier transform of $h[n]$, $H(\omega)$ is essentially the continuous-time Fourier transform of $h(t)$ restricted between $-\Omega_m$ and Ω_m and then normalised.

6.15.4. Graphical view of restriction and normalisation of $H(\Omega)$

Restriction

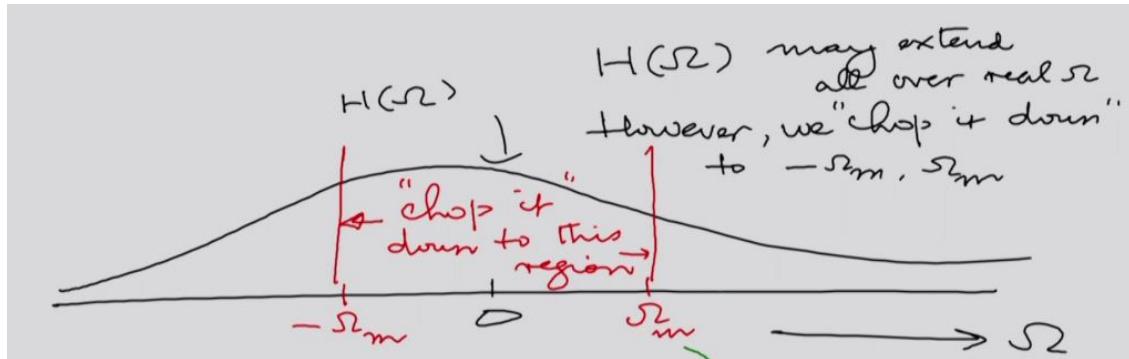


Figure 6.56.: Restriction on the frequency axis

Normalisation

Let us sample the signal at the frequency Ω_s such that $\Omega_s > 2\Omega_m$. Hence instead of restricting $H(\Omega)$ only to Ω_m , it would be better if we restrict it to $\Omega_s/2$, in order to include many more band-limited signals. Hence we normalise $\Omega_s/2$ to π and $-\Omega_s/2$ to $-\pi$. Now we obtain the discrete-time Fourier transform $H(\omega)$ of the desired

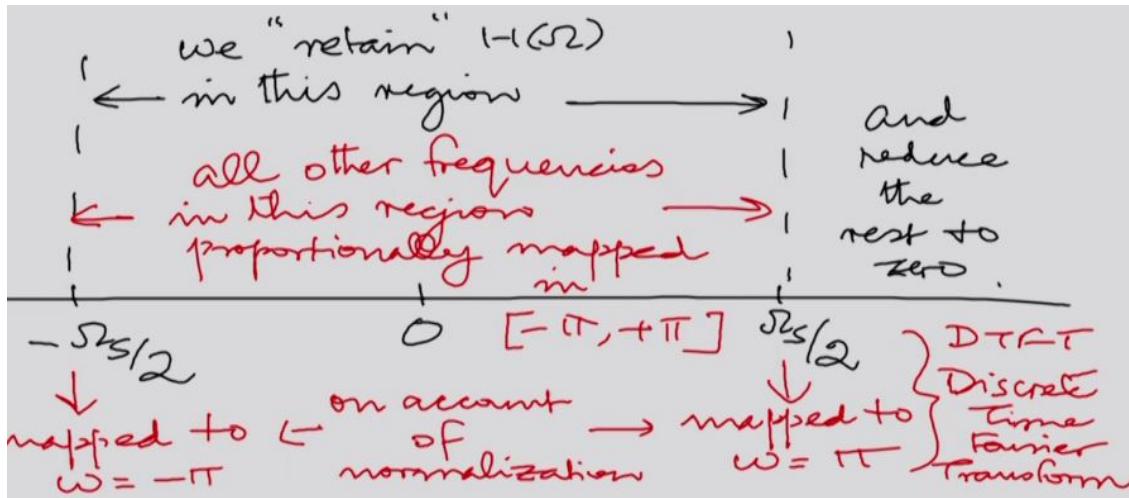


Figure 6.57.: Normalisation of the frequency axis with respect to sampling frequency

discrete-time system.

6. chapter

The whole process of DSP of the equivalent continuous-time counterparts, can now be summarised in Fig.6.58.

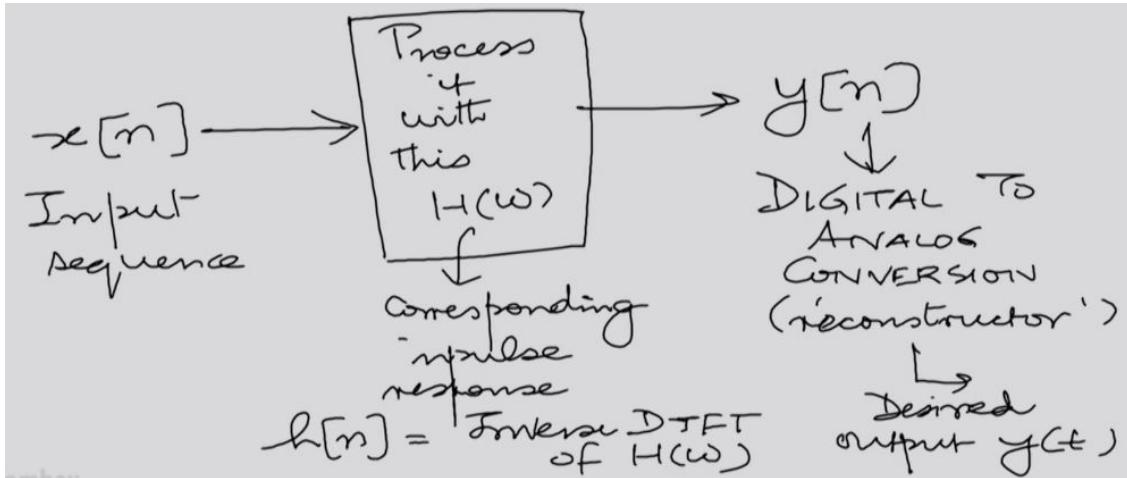


Figure 6.58.: Digital equivalent of the continuous-time signal processing

6.15.5. Example of Constructing a Discrete-Time System Equivalent to an RC Filter

Let the problem be as shown in Fig.6.59. We want to pass a signal band-limited to

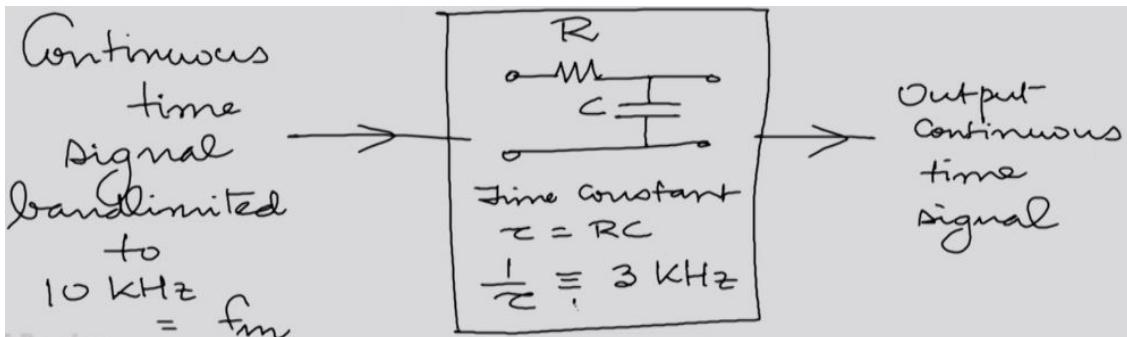


Figure 6.59.: Continuous-time processing that we wish to do

10kHz through an RC filter (low-pass) with a time constant 'equivalent' to 3kHz.

In order to find the equivalent discrete-time system, we need to know the frequency response of the continuous-time system. So let us find the frequency response of the system, the RC filter given in Fig.6.60. The impedance of the capacitor at a frequency of Ω is $\frac{1}{j\Omega C}$. From voltage divider principle,

$$\begin{aligned} \frac{V_{out}(\Omega)}{V_{in}(\Omega)} &= \frac{\frac{1}{j\Omega C}}{R + \frac{1}{j\Omega C}} \\ &= \frac{1}{1 + jRC\Omega} \\ &= \frac{1}{1 + j\tau\Omega} \end{aligned} \tag{6.8}$$

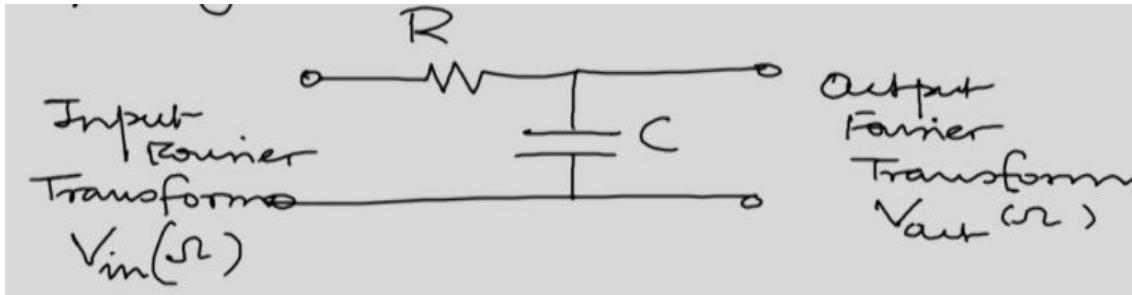


Figure 6.60.: Frequency response of RC filter

The frequency response has a magnitude of $\frac{1}{\sqrt{1+\tau^2\Omega^2}}$.

6.15.6. Significance of certain frequencies

Half-power frequency

At $\Omega = \frac{1}{\tau}$, the magnitude of the ratio of output to input becomes

$$\begin{aligned} |H\left(\frac{1}{\tau}\right)| &= \frac{1}{\sqrt{1+1}} \\ &= \frac{1}{\sqrt{2}} \end{aligned} \tag{6.9}$$

Power is proportional to the voltage squared. Hence the power in the output sinusoid at angular frequency $\frac{1}{\tau}$ is *half* that of the power in the input sinusoid.

K-powerpoint frequency

The frequency at which the power falls to K times the input power is given by,

$$\begin{aligned} |H(\Omega)|^2 &= \frac{1}{1+\tau^2\Omega^2} = K \\ \frac{1}{1+\tau^2\Omega^2} &= K \\ 1+\tau^2\Omega^2 &= \frac{1}{K} \\ \Omega^2 &= \frac{1}{\tau^2} \left(\frac{1}{K} - 1 \right) \\ \Omega &= \sqrt{\frac{1}{\tau^2} \left(\frac{1}{K} - 1 \right)} \\ \Omega &= \frac{1}{\tau} \sqrt{\left(\frac{1}{K} - 1 \right)} \end{aligned} \tag{6.10}$$

If K is 0.1, we fall to 10% of the input power. The frequency at which this

happens,

$$\begin{aligned}\Omega &= \frac{1}{\tau} \sqrt{\left(\frac{1}{0.1} - 1\right)} \\ &= \frac{3}{\tau}\end{aligned}\quad (6.11)$$

This is why we said earlier that $\frac{1}{\tau}$ is 'equivalent' to 3kHz. It has to be equal to the corresponding angular frequency and not the cycles/second frequency. Therefore, $\frac{1}{\tau} = 2\pi 3$ kHz. Therefore, the RC filter is a low-pass filter where at 9kHz (cycles/sec frequency), the power falls to 10%; at 3kHz, it falls to 50% and so on.

6.15.7. Description of the equivalent discrete-time RC filtering process

As the input signal is band-limited to 10kHz, we have to sample at a frequency greater than 20kHz, say 24kHz. From normalisation, we have that $2\pi 12$ kHz corresponds to π on the ω axis. Hence we need to consider the frequency response of the RC-filter also only upto $2\pi 12$ kHz as shown in Fig. 6.61. So the frequency of

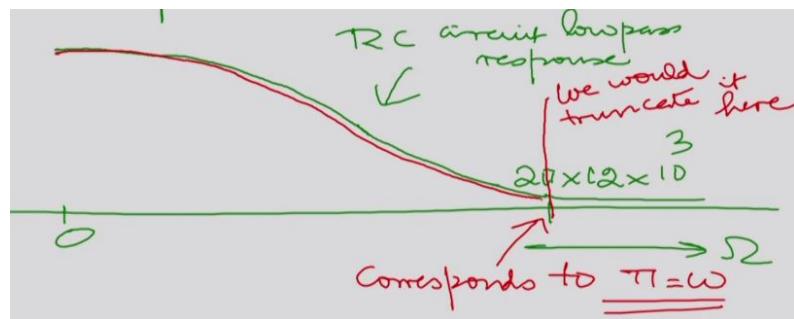


Figure 6.61.: Truncated frequency response to obtain the discrete-time frequency response

the discrete-time filter would be,

$$H(\omega) = \frac{1}{1 + j\tau\Omega} \quad (6.12)$$

with the Ω normalised as following. As $\frac{\Omega_s}{2}$ corresponds to $\omega = \pi$, Ω corresponds to,

$$\begin{aligned}\omega &= \Omega \frac{2\pi}{\Omega_s} \\ &= \frac{\Omega}{f_s} \\ &= \Omega T_s\end{aligned}\quad (6.13)$$

Therefore, $\Omega = \frac{\omega}{T_s}$. Hence,

$$H(\omega) = \frac{1}{1 + j\tau \frac{\omega}{T_s}} \quad (6.14)$$

After normalisation, we have to restrict ω between $-\pi$ and π . Also recall that since $H(\omega)$ is a discrete-time Fourier transform, it is periodic with period 2π . Therefore, periodically repeat the frequency response to obtain the full frequency response of

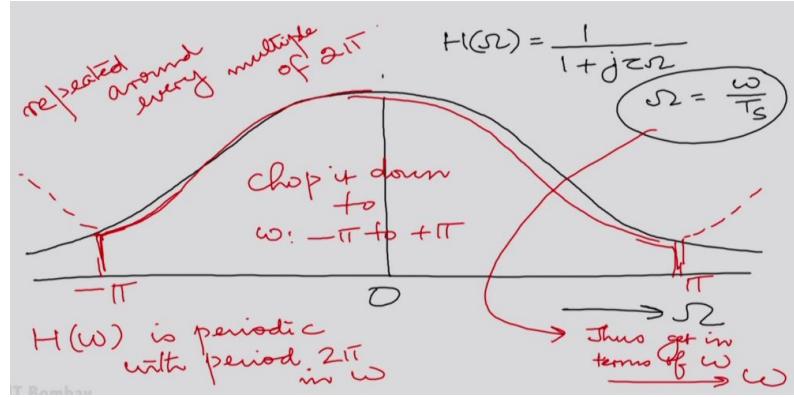


Figure 6.62.: Normalised, restricted and periodically extrapolate the continuous-time frequency response to obtain the discrete-time frequency response

the desired discrete-time system. We can, in principle, find the impulse response $h[n]$ by taking the inverse DTFT of $H(\omega)$, which would be complicated. We can, on the other hand, make an approximation. We could think of the filter as an ideal low-pass filter where $\Omega = \frac{3}{\tau} = 9$ kHz is in the stop-band and $\Omega = \frac{1}{\tau} = 3$ kHz is in the pass-band. We can choose the cut-off frequency to be somewhere in between (say, 6kHz) and treat it as an almost ideal low-pass filter for convenience of calculations.

6.16. Lecture 16: Conclusion of Module 3

6.16.1. Introduction

In this lecture we will look at simple replacement of continuous time RC circuit system with its equivalent discrete system. And also understand the purpose behind wanting to process signals digitally instead of in analog.

6.16.2. Recap

In the previous lecture we looked at an example of equivalent discrete time processing system for an RC circuit. The system we had looked at has the following feature $\frac{1}{\tau} \equiv 3\text{kHz}$ is the half power point and $\frac{1}{\tau} \equiv 9\text{kHz}$ is the 10% power point.

6.16.3. Construction of equivalent discrete system

The DTFT of the discrete equivalent system $H(\omega)$ is essentially Fourier transform $H(\Omega)$ restricted to $-\Omega_m$ to $+\Omega_m$. We could take exact chopped(restricted) system of the RC response itself $-\pi$ to $+\pi$ treat that as frequency response of discrete system.

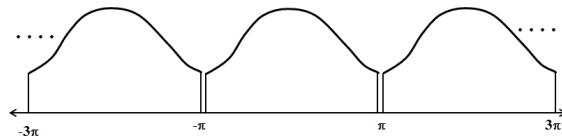


Figure 1:DTFT of discrete equivalent system

6. chapter

Or we could choose ideal low pass filter which behaves very similar to RC response.

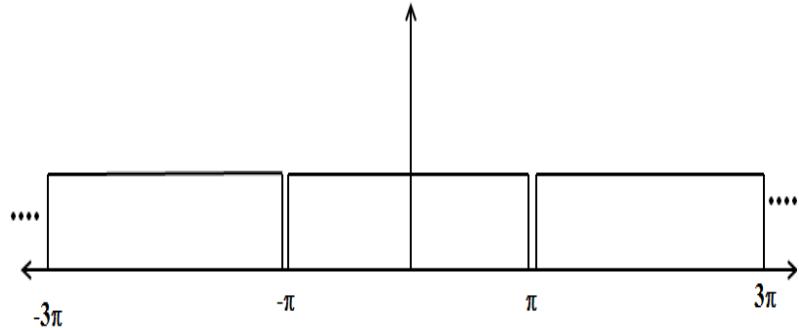


Figure 2: Ideal low pass filter similar to $H(\omega)$

The second way is of course simpler. Low pass filter is characterised by cut off frequency. We can suitably choose the cut off at some frequency between 3 kHz and 9 kHz depending on where we are satisfied. We could choose 6 kHz as cutoff for ideal low pass filter.

Ideal low pass filter with cut off 6kHz is equivalent to discrete time system with the following frequency response.

$$\omega = \Omega T_s$$

Let us consider $T_s = \frac{1}{24\text{kHz}}$

$$\omega = \frac{6}{24} 2\pi = \frac{\pi}{2}$$

Ideal low pass filter on ω scale

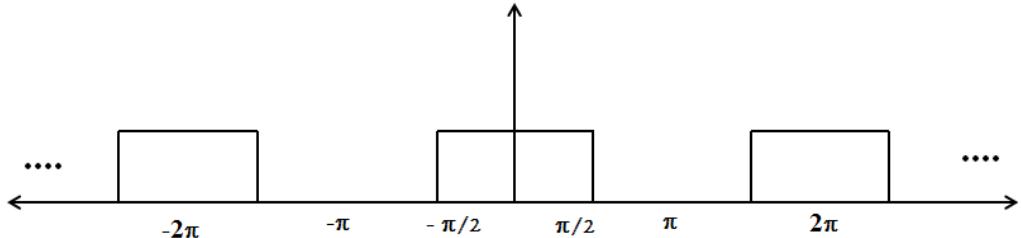


Figure 3: Ideal low pass filter with cut off 6 kHz on ω scale

$H(\omega)$ is DTFT of corresponding impulse response $h[n]$.

$$H(\omega) = \sum_{n=-\infty}^{+\infty} h[n] e^{-j\omega n}$$

Note this is periodic with period 2π

$$H(\omega + 2\pi) = \sum_{n=-\infty}^{+\infty} h[n] e^{-j(\omega+2\pi)n}$$

$$H(\omega + 2\pi) = \sum_{n=-\infty}^{+\infty} h[n] e^{-j\omega n} e^{-j2\pi n}$$

Note $e^{-j2\pi n} = 1$ for all integer n and therefore the periodicity.

6.16.4. Inverse Discrete Time Fourier Transform

Think $H(\omega)$ as the Fourier transform of a continuous time impulse response sampled at unit frequency.

Corresponding continuous time impulse response in the inverse Fourier transform

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} H(\omega) e^{j\omega t} d\omega$$

We can get impulse response of the discrete system by replacing $t = n$. Impulse response of DT system is given by inverse DTFT of $H(\omega)$,

$$h[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} H(\omega) e^{j\omega n} d\omega$$

Exercise: Let $H(\omega) = 1$ for $-\frac{\pi}{2} \leq \omega \leq \frac{\pi}{2}$, else 0.

$$h[n] = \frac{1}{2\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} 1 \cdot e^{j\omega n} d\omega$$

Evaluating separately for $n = 0$ and $n \neq 0$.

$$h[n] = \frac{1}{2\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} 1 \cdot d\omega, n = 0$$

$$= \frac{1}{2\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} 1 \cdot e^{j\omega n} d\omega, n \neq 0$$

After simple calculations we get the expression for $h[n]$ as

$$\begin{aligned} h[n] &= \frac{1}{2}, n = 0 \\ &= \frac{\sin \frac{\pi}{2} n}{\pi n}, n \neq 0 \end{aligned}$$

Keep in mind that we oversimplified the RC circuit response. It is possible to find out the exact impulse response of the RC circuit chopped frequency response.

6.16.5. Exact Impulse Response

$$H(\Omega) = \frac{1}{1 + j\Omega\tau}$$

$$\Omega = \frac{\omega}{T_s}$$

$$H(\omega) = \frac{1}{1 + j\omega \frac{\tau}{T_s}} : -\pi \leq \omega \leq \pi$$

6. chapter

Recall we restricted the frequency response from $-\pi$ to $+\pi$ exact $h[n] = \text{Inverse DTFT of } H(\omega)$.

$$h[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{j\omega n}}{1 + j\omega \frac{T}{T_s}} d\omega$$

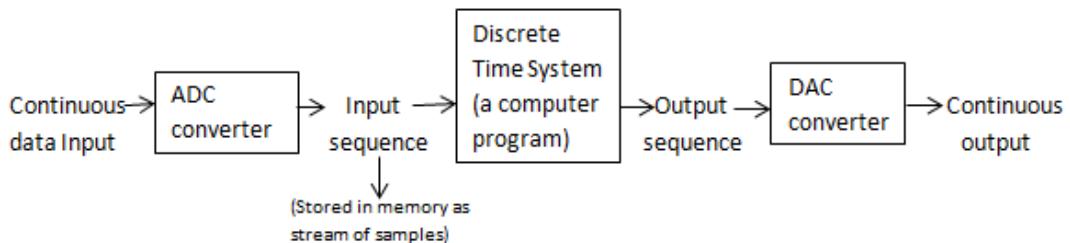
Calculate $h[n]$ and compare with low pass filter that we have taken.

6.16.6. Conclusion of this module

Continuous systems are found everywhere in real life. Discrete systems occur sometimes and therefore continuous independent variable make lot of sense if you want to understand real life situation. In general it is continuous variable that dominates seemingly in the word around us.

We could of-course construct artificial discrete time system. Economic and financial are good examples. All systems that operate at certain points of the clock are good examples.

But continuous independent variable is much more difficult to understand and process. The convenience of processing discrete independent variable comes from the fact that all the discrete time operation can be done by storing the input samples in a computer, processing them on a computer with a discrete time system which is implemented in the form of a small program. The moment you are in discrete time you can program it by having a loop that repeats at every instant n . A discrete time system can be implemented with modern computing devices. In a discrete system for an input sequence you can write a computer program to get the output sequence.



Module 3 has given you the tools to interchange between discrete independent variable and continuous independent variable. By this time you must have noted that continuous independent time is reality and discrete independent time is convenience.

Part IV.

Laplace and Z-Transforms

7. chapter

7.1. Module 4: Lecture 1 Introduction to Z Transform

7.1.1. Introduction

We shifted from natural domain to frequency domain to analyse more clearly what system does to a signal. However, in the previous module, we considered only those *stable functions* whose Fourier transform exists. We need to generalise the results obtained from the Fourier domain analysis to a more general domain. In this module we will be studying the following things:

1. Generalising the Fourier transform to accommodate systems whose impulse response do not have a Fourier transform. For example,

Continuous time domain example: $h(t) = e^t u(t)$

Discrete time domain example: $h(t) = 2^n u[n]$

We know that $\sum_{n=0}^{\infty} 2^n$ is divergent which makes the system unstable. The above two examples of unstable system have no frequency response.

2. Generalising the properties of Fourier Transform to a more general context, like,

- Convolution
- Shift Invariance
- Differentiation
- Modulation

3. Characterizing the system more generally to deal with more general classes of inputs in an elegant manner.

We require a more general transformation framework which can basically capture the source of instability of the signal and then apply the Fourier transform on the modified function.

For example, $h(t) = e^t u(t)$ is an unstable function with no Fourier transform because of exponential growth of e^t . We can capture this source of instability by multiplying the function by a decaying exponential stronger than the growing one. That is, $h(t)e^{-\sigma t}$ for $\sigma \geq 1$ results in $h(t)e^{-\sigma t} = e^{(1-\sigma)t} u(t)$, which is absolutely integrable.

For a continuous independent variable, this whole process is known as the *Laplace transform*, and for a discrete independent variable, it is called the *Z transform*.

7.1.2. Introduction of the Z Transform

The process of capturing or holding down a unstable discrete time independent signal, and then taking its discrete time Fourier transform is what taking a Z transform entails. For example, consider $x[n] = 2^n u[n]$, i.e.

$$x[n] = 2^n \text{ for } n \geq 0$$

$$x[n] = 0 \text{ for } n < 0$$

Clearly this sequence doesn't have a discrete time Fourier transform. We can capture the growth of this signal by multiplying $x[n]$ by r^{-n} . The modified sequence is $2^n r^{-n}$, where r is a *capturing factor*. For a discrete time Fourier Transform of $x[n]r^{-n}$ to exist, $\sum_{n=0}^{\infty} x[n]r^{-n}e^{-jwn}$ should be finite. Here,

$$\sum_{n=0}^{\infty} x[n]r^{-n}e^{-jwn} = \sum_{n=0}^{\infty} (2r^{-1}e^{-jw})^n$$

For this geometric series to converge, the common ratio $(2r^{-1}e^{-jw})$ has to be less than 1. i.e. we have,

$$\begin{aligned} |2r^{-1}e^{-jw}| &< 1 \\ \implies |2r^{-1}| &< 1 \\ \implies |r| &> 2 \end{aligned}$$

We can also write $z = re^{-jw}$, where z is a complex number. Hence, $|r| > 2 \implies |z| > 2$ which can be represented in the complex plane by the region exterior to the circle $|z| = 2$.

This region exterior to the circle, where the sequence converges, is called the *region of convergence*. Hence, now we have a more general transform on the sequences. We call this the Z transform of the sequence.

$$X(z) = \sum_{n=0}^{\infty} x[n]z^{-n}$$

In the above example,

$$X(z) = \sum_{n=0}^{\infty} (2r^{-1}e^{-jw})^n = \sum_{n=0}^{\infty} 2^n z^{-n}$$

Therefore,

$$X(z) = \frac{1}{1 - 2z^{-1}}$$

is the Z - Transform of $x[n]$, provided $|z| > 2$. To state the Z transform, we need to specify both the expression *and* the region of convergence of the expression.

To illustrate this, consider,

$$x[n] = -(2)^n \text{ for } n \leq 1$$

$$x[n] = 0 \text{ elsewhere}$$

Now, the Z Transform $X(z)$ is given by,

$$X_1(z) = \sum_{n=-\infty}^1 -2^n z^{-n} = 2^{-1}z^1 + 2^{-2}z^2 + \dots$$

It is a geometric series with common ratio $2^{-1}z$. For this series to converge, the common ratio has to be less than 1, i.e. we have,

$$|2^{-1}z| < 1$$

which implies

$$|z| < 2$$

Therefore, we have

$$X_1(z) = -\frac{2^{-1}z}{1 - 2^{-1}z} = -\frac{-1}{2z^{-1} - 1} = \frac{1}{1 - 2z^{-1}}$$

with the Region of Convergence $|z| < 2$. Clearly we can see that $X(z)$ and $X_1(z)$ have the *same* expression but *different* Regions of Convergence. This explains the importance of Region of Convergence in Z Transform.

7.2. Module 4: Lecture 2

Defining the Laplace Transform and the Z Transform

7.2.1. Introduction

In the previous lectures you were introduced to the basic idea of extending the definition of the Fourier Transform to capture the unstable signals. It also exposed you to the definition of the *Z transform* for Discrete Time Fourier Transform. In this lecture, you will be introduced to *Laplace Transform* on similar motivation of the Continuous case.

Also, throughout the subsequent lecture we will follow the convention that the Laplace transform of $x(t)$ will be denoted $X(s)$, and the region of convergence by R_X . And this relation is shown by

Laplace transform of $x(t)$: $x(t) \xrightarrow{\mathcal{L}} X(s) ; R_X$.

Also, Fourier transform of $x(t)$: $x(t) \xrightarrow{\mathcal{F}} X(j\Omega) ; R_X$.

and Z-transform of $x(t)$: $x[n] \xrightarrow{z} X(z) ; R_X$

7.2.2. Laplace Transform

Consider the input $x(t) = e^t u(t)$. Notice that this is an exponentially growing unstable signal. As per our earlier discussion, we need to capture this growing sequence and take its Fourier Transform. For this, we will need to multiply it by a decaying exponential, let's say $e^{-\sigma t}$. Now, if we take the Fourier Transform of $x(t)e^{-\sigma t}$

$$\begin{aligned} X(\sigma + j\Omega) &= \int_{-\infty}^{\infty} x(t)e^{-\sigma t}e^{-j\Omega t} dt \\ &= \int_0^{\infty} e^t e^{-\sigma t}e^{-j\Omega t} dt \end{aligned}$$

7. chapter

$$= \int_0^\infty e^{(1-\sigma)t} e^{-j\Omega t} dt$$

Asking when the above Fourier Transform $X(s)$ will exist, is equivalent to asking when will the above integral converge and what are the conditions on σ and Ω .

Note that it will only be the $e^{(1-\sigma)t}$ which would contribute to the convergence as the other factor $e^{-j\Omega t}$ has magnitude 1 and contributes only in the phase part. Thus, $e^{(1-\sigma)t}$ needs to be a decaying exponential. It should be a negative power of e which would happen if $(1 - \sigma) < 0$ which implies that $\sigma > 1$. If we consider $\sigma + j\Omega = s$ where s is a complex constant with $\Re(s) = \sigma$ and $\Im(s) = \Omega$ then the Fourier Transform becomes

$$X(s) = \int_{-\infty}^\infty x(t) e^{-st} dt$$

In our considered example (for $\sigma > 1$ or $\Re(s) > 1$),

$$\begin{aligned} X(s) &= \int_0^\infty e^{(1-\sigma)t} e^{-j\Omega t} dt \\ &= \int_0^\infty e^t e^{-st} dt \\ &= \int_0^\infty e^{(1-s)t} dt \\ &= \frac{1}{(1-s)} e^{(1-s)t} \Big|_0^\infty \\ \mathbf{X}(s) &= \frac{\mathbf{1}}{(s-1)} \end{aligned}$$

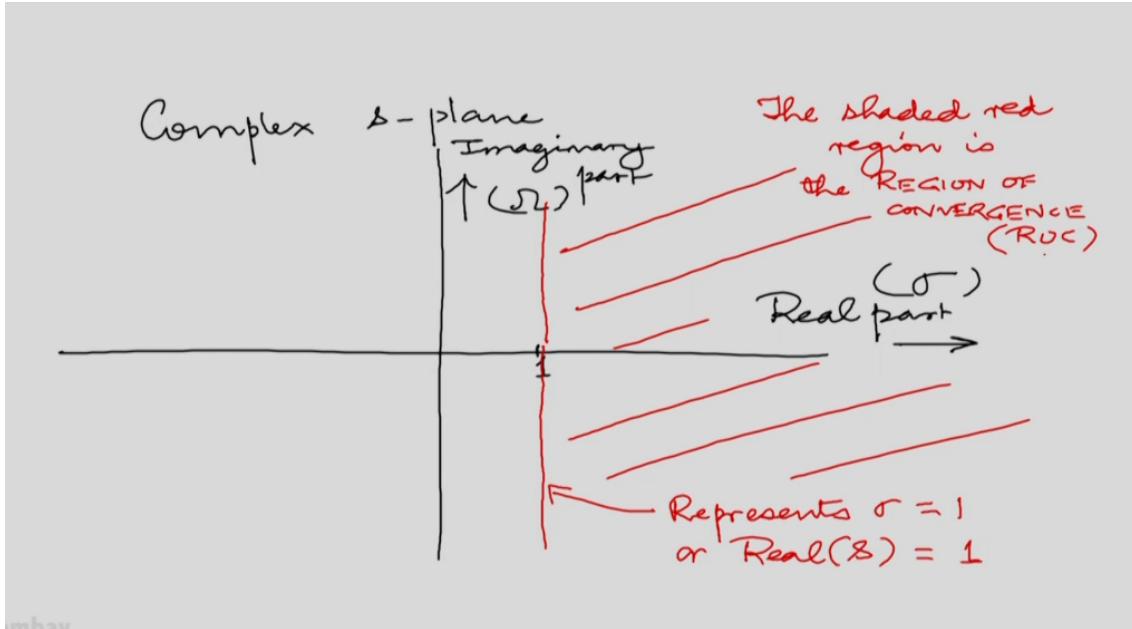
The region in the complex s plane in which the above formulation exists is known as the Region Of Convergence of the *Laplace Transform* which can be graphically shown by the figure below.

Now consider $x_1(t) = -e^t u(-t)$ which is the same exponential signal on the negative half in time domain.

$$\begin{aligned} X_1(s) &= \int_{-\infty}^\infty x_1(t) e^{-st} dt \\ &= \int_{-\infty}^0 -e^t e^{-st} dt \\ &= \int_{-\infty}^0 -e^{(1-s)t} dt \\ &= \frac{-1}{(1-s)} e^{(1-s)t} \Big|_{-\infty}^0 \\ \mathbf{X}_1(s) &= \frac{\mathbf{1}}{(s-1)} \end{aligned}$$

But observe that during the second last step we need to realise that for the definite integral to exist, we need $e^{(1-s)t}$ to go down to 0 as t tends to $-\infty$. Thus we have $\Re(1 - s) > 0$ which implies $\Re(s) < 1$.

Thus, Laplace Transform of $x_1(t) = -e^t u(-t)$ is $\mathbf{X}_1(s) = \frac{\mathbf{1}}{(s-1)}$ with $R_X : \Re(s) < 1$ or $\sigma < 1$. It is exactly the complement graph of that shown in Figure 1.

Figure 7.1.: Region of convergence R_X for the input $x(t) = e^t u(t)$

7.2.3. Significance of the Region of Convergence

We have seen that the *expression* for the Laplace transform of $x(t)$ and $x_1(t)$ come out to be exactly the same, even though they are different signals. This indicates that the Laplace Transform is not only the expression in terms of s but also the Region of Convergence(ROC). If we don't consider ROC also, we might be looking at a wrong expression!

$$e^t u(t) \xrightarrow{\mathcal{L}} \frac{1}{(s-1)} ; \text{Re}(s) > 1.$$

$$-e^t u(-t) \xrightarrow{\mathcal{L}} \frac{1}{(s-1)} ; \text{Re}(s) < 1.$$

We start by observing that the Laplace Transform of both the signals is the same and the only difference is the ROC which are mutually exclusive regions. For a signal to apriori have a Fourier Transform we need to have $\Re(s) = 0$ i.e. $\sigma = 0$ contained in the ROC. This is because if $\sigma = 0$, then $s = j\Omega$, and the expression for the Laplace transform just reduces to that of the Fourier transform. Thus, if the Laplace transform does not exist for $\sigma = 0$, that means that the Fourier transform too does not exist. Thus we conclude that the signal with $\Re(s) = 0$ in its region of convergence does a priori have a Fourier Transform but for the other, we will have to hold it down so that the Fourier Transform exists. In our example, $x_1(t)$ contains $\Re(s) = 0$ in its ROC, and hence has a Fourier transform, whereas $x(t)$ doesn't. To have a comparative picture of the two discussed signals, observe the Figure 2.

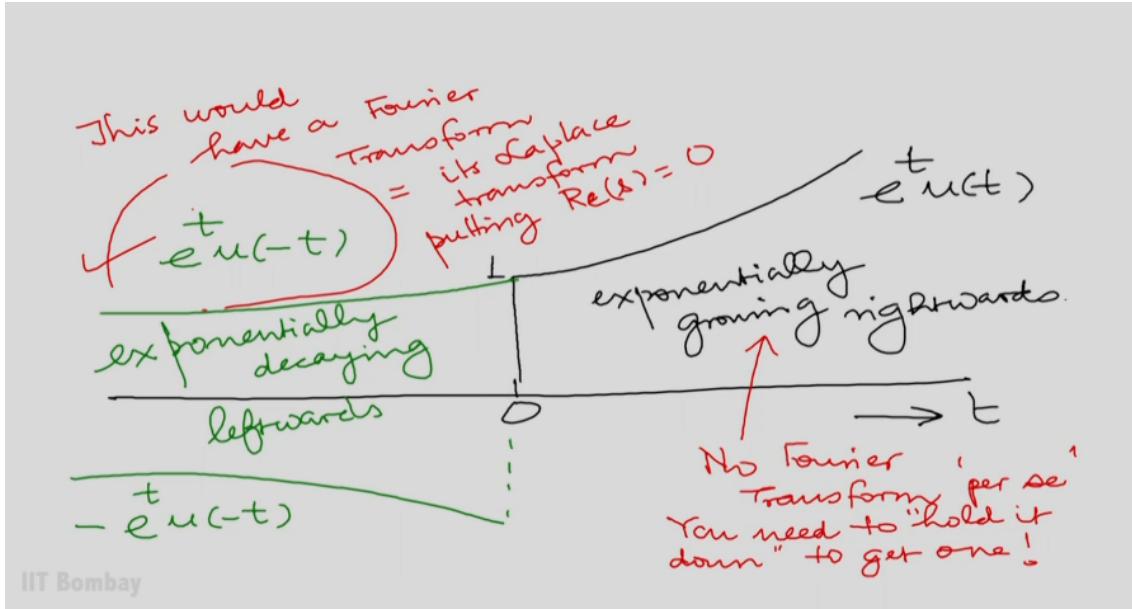


Figure 7.2.: Region of convergence R_X for the input $x(t) = e^t u(t)$

7.2.4. Z Transform Equivalent Summary

The Z-transform of a sequence $x[n]$ is given by $X(z)$

$$X(z) = \sum_{n=-\infty}^{\infty} x[n]z^{-n}$$

For example if $x[n] = 3^n u[n]$

$$X(z) = \sum_{n=-\infty}^{\infty} 3^n u[n]z^{-n}$$

$$X(z) = \sum_{n=0}^{\infty} \left(\frac{3}{z}\right)^n$$

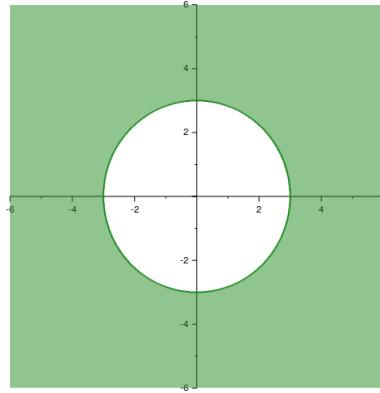
$$\mathbf{X(z)} = \frac{\mathbf{z}}{\mathbf{z}-3}$$

With the Region of Convergence :

$$\mathbf{R_X : |z| > 3}$$

7.2.5. Conclusion

In this lecture, we looked at the definition of *Laplace transform* with some basic examples. It also invoked the importance of the Region of Convergence in Laplace Transforms and its equivalence in Z-Transforms. In the subsequent lectures we will go about solving some examples of both Laplace and Z-transforms.

Figure 7.3.: Region of convergence R_X for the input $x[n] = 3^n u[n]$

7.3. Module 4: Lecture 3

Examples of Laplace and Z Transforms

7.3.1. Introduction

In the previous lectures you were introduced to the definition of the *Laplace transform* and the *z transform*. In this lecture, we will look at a few examples of these transforms.

Also, throughout the subsequent lectures we will follow the convention that $x(t)$ has the Laplace transform $X(s)$, in the region of convergence R_X . And this relation is shown by

$$x(t) \xrightarrow{\mathcal{L}} X(s), \quad R_X$$

Similarly for the input $x[n]$ and its z-transform $X(z)$, in the region of convergence R_X , we write

$$x[n] \xrightarrow{z} X(z), \quad R_X$$

7.3.2. Laplace Transform Example

Here, we will assume the input $x(t) = e^{2t}u(t) + e^{3t}u(t)$. We wish to find its *Laplace transform* and the region of convergence.

$$\begin{aligned} X(s) &= \int_{-\infty}^{\infty} x(t)e^{-st}dt \\ &= \int_{-\infty}^{\infty} (e^{2t}u(t) + e^{3t}u(t))e^{-st}dt \\ &= \int_{-\infty}^{\infty} e^{2t}u(t)e^{-st}dt + \int_{-\infty}^{\infty} e^{3t}u(t)e^{-st}dt \\ &= \int_0^{\infty} e^{(2-s)t}dt + \int_0^{\infty} e^{(3-s)t}dt \end{aligned}$$

Now the left term will converge only if $(2 - Re(s)) < 0$, i.e., $Re(s) > 2$. Similarly, the right term will converge only if $Re(s) > 3$. Hence the region of convergence is

$$\begin{aligned} R_X &= (Re(s) > 2) \cap (Re(s) > 3) \\ &= Re(s) > 3 \end{aligned}$$

7. chapter

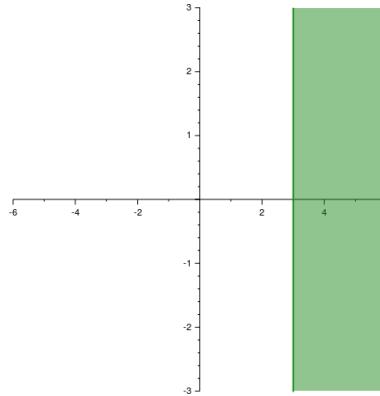


Figure 7.4.: Region of convergence R_X for the input $x(t) = e^{2t}u(t) + e^{3t}u(t)$. (The green shaded region represents the ROC $\text{Re}(s) > 3$)

Continuing on the integral,

$$\begin{aligned} X(s) &= \frac{1}{(2-s)}e^{(2-s)t}\Big|_0^\infty + \frac{1}{(3-s)}e^{(3-s)t}\Big|_0^\infty \\ &= \frac{1}{(s-2)} + \frac{1}{(s-3)} \end{aligned}$$

7.3.3. Z Transform Example

Here, we will assume the input $x[n] = 2^n u[n] + 3^n u[n]$. We wish to find its z -transform and the region of convergence.

$$\begin{aligned} X(z) &= \sum_{n=-\infty}^{\infty} x[n]z^{-n} \\ &= \sum_{n=-\infty}^{\infty} (2^n u[n] + 3^n u[n])z^{-n} \\ &= \sum_{n=-\infty}^{\infty} 2^n u[n]z^{-n} + \sum_{n=-\infty}^{\infty} 3^n u[n]z^{-n} \\ &= \sum_{n=0}^{\infty} \left(\frac{2}{z}\right)^n + \sum_{n=0}^{\infty} \left(\frac{3}{z}\right)^n \end{aligned}$$

Here, the left term only converges if $\left|\frac{2}{z}\right| < 1$, i.e., $|z| > 2$. Similarly from the right term, $|z| > 3$. Again, the region of converge

$$\mathbf{R}_X : |z| > 3$$

Continuing on the summation,

$$X(z) = \frac{z}{z-2} + \frac{z}{z-3}$$

7.3.4. Conclusion

In this lecture, we looked at a few examples of the *Laplace transform* and the *z-transform*. In the subsequent lectures, we will be looking at some properties of the *Laplace transform* and the *z-transform*, like linearity, modulation property, etc.

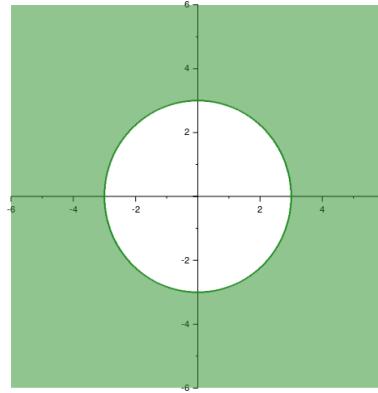


Figure 7.5.: Region of convergence R_X for the input $x[n] = 2^n u[n] + 3^n u[n]$. (The green shaded region represents the ROC $|z| > 3$)

7.4. Module 4: Lecture 4 Properties of Laplace and Z Transforms

7.4.1. Introduction

In the previous lectures you were introduced to Laplace transforms and z-transforms. In this lecture, we will discuss some important properties of these transforms, like the linearity, shifting and modulation property.

Also, throughout the subsequent lectures we will follow the convention that $x(t)$ has the Laplace transform $X(s)$, in the region of convergence R_X . And this relation is shown by

$$x(t) \xrightarrow{\mathcal{L}} X(s), R_X$$

Similarly for the input $x[n]$ and its z-transform $X(z)$, in the region of convergence R_X , we write

$$x[n] \xrightarrow{z} X(z), R_X$$

7.4.2. Linearity

Laplace Transform

Property

The linearity property of the Laplace transform states that if,

$$\begin{aligned} x_1(t) &\xrightarrow{\mathcal{L}} X_1(s), R_{X_1} \\ x_2(t) &\xrightarrow{\mathcal{L}} X_2(s), R_{X_2} \end{aligned}$$

then the following property holds,

$$\alpha x_1(t) + \beta x_2(t) \xrightarrow{\mathcal{L}} \alpha X_1(s) + \beta X_2(s), R_X \supseteq R_{X_1} \cap R_{X_2}$$

7. chapter

Proof

For input $x(t) = \alpha x_1(t) + \beta x_2(t)$, we have

$$\begin{aligned} X(s) &= \int_{-\infty}^{\infty} (\alpha x_1(t) + \beta x_2(t)) e^{-st} dt \\ &= \alpha \int_{-\infty}^{\infty} x_1(t) e^{-st} dt + \beta \int_{-\infty}^{\infty} x_2(t) e^{-st} dt \\ &= \alpha X_1(s) + \beta X_2(s) \end{aligned}$$

Now, in the region given by $R_{X_1} \cap R_{X_2}$, we are assured that both the integrals are convergent, and hence $X(s)$ is convergent. However, we can see that s outside this region might also make $X(s)$ converge. Try the following example, to understand this case.

$$\begin{aligned} x_1(t) &= e^{2t} u(t) + e^{3t} u(t) \\ x_2(t) &= e^{2t} u(t) - e^{3t} u(t) \\ x(t) &= x_1(t) + x_2(t) \end{aligned}$$

You can clearly see that the regions of convergence for $X_1(s)$ and $X_2(s)$ is $\text{Re}(s) > 3$, but that of $X(s)$ is $\text{Re}(s) > 2$, which is a super-set of the intersection of the 2 regions.

Z-Transform

Property

A similar property and proof can be constructed for the z-transform. The linearity property of the z-transform states that if,

$$x_1[n] \xrightarrow{z} X_1(z), \quad R_{X_1}$$

$$x_2[n] \xrightarrow{z} X_2(z), \quad R_{X_2}$$

then the following property holds,

$$\alpha x_1[n] + \beta x_2[n] \xrightarrow{z} \alpha X_1(z) + \beta X_2(z), \quad R_X \supseteq R_{X_1} \cap R_{X_2}$$

Proof

For the input $x[n] = \alpha x_1[n] + \beta x_2[n]$

$$\begin{aligned} X(z) &= \sum_{n=-\infty}^{\infty} (\alpha x_1[n] + \beta x_2[n]) z^{-n} \\ &= \alpha \sum_{n=-\infty}^{\infty} x_1[n] z^{-n} + \beta \sum_{n=-\infty}^{\infty} x_2[n] z^{-n} \\ &= \alpha X_1(z) + \beta X_2(z) \end{aligned}$$

Once again, in the region given by $R_{X_1} \cap R_{X_2}$, we are assured that both the summations are convergent, and hence $X(z)$ is convergent. However, you can construct examples where z outside this region might also make $X(z)$ converge.

7.4.3. Shifting of Independent Variable

Laplace Transform

Property

The shifting property of the Laplace transform states that if,

$$x(t) \xrightarrow{\mathcal{L}} X(s), R_X$$

then the following property holds,

$$x(t - t_0) \xrightarrow{\mathcal{L}} e^{-st_0} X(s), R'_X$$

where R'_X is same as R_X except possibly for the extreme contours, $\text{Re}(s) \rightarrow \infty$ or $\text{Re}(s) \rightarrow -\infty$.

Proof

$$X'(s) = \int_{-\infty}^{\infty} x(t - t_0) e^{-st} dt$$

replacing $t - t_0$ with λ , and making corresponding transformation in integral,

$$X'(s) = \int_{-\infty}^{\infty} x(\lambda) e^{-s(\lambda+t_0)} d\lambda$$

$$X'(s) = e^{-st_0} X(s)$$

The region of convergence remains same, except for the extreme contours due to the multiplication of the e^{-st_0} term. The transform might not be defined or now become infinite for the extreme contours.

Example

You can see this property in action in the following example.

$$\delta(t) \xrightarrow{\mathcal{L}} 1, \text{entire } s \text{ plane}$$

Now, for the input

$$\delta(t - 1) \xrightarrow{\mathcal{L}} e^{-s}, R'_X$$

where R'_X is the entire s-plane excluding $\text{Re}(s) \rightarrow -\infty$, because the term e^{-s} becomes unbounded for $\text{Re}(s) \rightarrow -\infty$. You can try proving similar results for the input $\delta(t + 1)$.

Z-Transform

Property

The shifting property of the z-transform states that if,

$$x[n] \xrightarrow{z} X(z), R_X$$

then the following property holds,

$$x[n - n_0] \xrightarrow{z} z^{-n_0} X(z), R'_X$$

where R'_X is same as R_X except possibly for the extreme contours, $|z| \rightarrow 0$ or $|z| \rightarrow \infty$.

Proof

$$X'(z) = \sum_{n=-\infty}^{\infty} x[n - n_0] z^{-n}$$

Now, replacing $n - n_0$ with m and transforming the summation

$$X'(z) = \sum_{m=-\infty}^{\infty} x[m] z^{-(m+n_0)}$$

$$X'(z) = z^{-n_0} X(z)$$

Once again, we can argue that due to the term z^{-n_0} the extreme contours may be effected.

7.4.4. Modulation Property

Modulation refers to multiplying with fixed signal or sequence. Hence, here we will try to find the Laplace transform and z-transform for modulated signals.

Laplace Transform

Property

The modulation property of the Laplace transform states that if,

$$x(t) \xrightarrow{\mathcal{L}} X(s), R_X$$

then the following property holds,

$$e^{at}x(t) \xrightarrow{\mathcal{L}} X(s-a), R'_X$$

where a is a complex constant and R'_X is such that $(s-a)$ belongs to R_X .

Proof

$$\begin{aligned} X'(s) &= \int_{-\infty}^{\infty} e^{at}x(t)e^{-st}dt \\ &= \int_{-\infty}^{\infty} x(t)e^{-(s-a)t}dt \\ &= X(s-a) \end{aligned}$$

The region of convergence of $X'(s)$ is the same as that of $X(s-a)$. Hence it is the same as $(s-a)$ lying in R_X .

Example

For the input $x(t) = e^{2t}u(t)$ and $a = 3$,

$$e^{3t}x(t) \xrightarrow{\mathcal{L}} \frac{1}{(s-3)-2} = \frac{1}{s-5}$$

$$R'_X = Re(s-3) > 2 = Re(s) > 5$$

solving directly, $x'(t) = e^{3t}e^{2t}u(t) = e^{5t}u(t)$, for which

$$e^{5t}u(t) \xrightarrow{\mathcal{L}} \frac{1}{s-5}, \text{Re}(s) > 5$$

which is what we got from the property.

Z-Transform

Property

The modulation property of the z-transform states that if,

$$x[n] \xrightarrow{z} X(z), R_X$$

then the following property holds,

$$\alpha^n x[n] \xrightarrow{z} X\left(\frac{z}{\alpha}\right), R'_X$$

where α is a complex constant and R'_X is such that $\frac{z}{\alpha}$ belongs to R_X .

Proof

$$\begin{aligned} X'(z) &= \sum_{n=-\infty}^{\infty} \alpha^n x[n] z^{-n} \\ &= \sum_{n=-\infty}^{\infty} x[n] \left(\frac{z}{\alpha}\right)^{-n} \\ &= X\left(\frac{z}{\alpha}\right) \end{aligned}$$

Once again, the region of convergence of $X'(z)$ is the same as that of $X\left(\frac{z}{\alpha}\right)$. Hence it is the same as $\frac{z}{\alpha}$ lying in R_X .

Exercise

For the input $x[n] = 2^n u[n]$ and $\alpha = \frac{1}{3}$, try proving that the modulated sequence has the z-transform $\frac{1}{1 - \frac{2}{3}z^{-1}}$, and region of convergence $|z| > \frac{2}{3}$, using the modulation property.

7.4.5. Introduction To Convolution Property

Here we will be introducing the convolution property of Laplace transforms. In the subsequent lecture, we will be discussing this in more detail and completeness and extend it to sequences and z-transform.

Laplace Transforms

Property

For the signals $x(t)$ and $h(t)$ say,

$$x(t) \xrightarrow{\mathcal{L}} X(s), R_X$$

$$h(t) \xrightarrow{\mathcal{L}} H(s), R_H$$

The convolution property states that for the input $y(t) = x(t) * h(t)$, the Laplace transform is given by

$$Y(s) = X(s)H(s), R_Y$$

Proof

$$\begin{aligned} Y(S) &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau \right] e^{-st} dt \\ Y(S) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(\tau)h(t-\tau)e^{-st} d\tau dt \end{aligned}$$

Now, replace $t - \tau$ with λ for fixed τ .

$$\begin{aligned} Y(S) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(\tau)h(\lambda)e^{-s(\lambda+\tau)} d\tau d\lambda \\ &= \left[\int_{-\infty}^{\infty} x(\tau)e^{-s\tau} d\tau \right] \left[\int_{-\infty}^{\infty} h(\lambda)e^{-s\lambda} d\lambda \right] \\ &= X(s)H(s) \end{aligned}$$

7.4.6. Conclusion

In this lecture you were introduced to linearity, shifting, modulation and convolution properties. We will further carry on the discussion on the convolution property in the subsequent lectures.

7.5. Module 4: Lecture 5

Region of Convergence(ROC), Convolution Property and System Function

7.5.1. Introduction

In the last session we saw the very important property of the Laplace transform and z-transform that, when we convolve two signals, their Laplace transform or z-transform are multiplied. In this session we talk about region of convergence(ROC) of the Laplace transform and z-transform of output and how Laplace transform or z-transform of the impulse response together with its ROC gives complete information about the system.

We saw that if

$$x(t) \xrightarrow{\mathcal{L}} X(s); R_X$$

and

$$h(t) \xrightarrow{\mathcal{L}} H(s); R_H$$

then if $y(t) = x(t) * h(t)$, then

$$y(t) \xrightarrow{\mathcal{L}} X(s)H(s) ; R_Y$$

This is called the convolution property of the Laplace transform.

What will be the ROC of $Y(s)$, and how will it be related to the ROCs of $X(s)$ and $H(s)$?

7.5.2. ROC

We saw that the Laplace transform of $y(t)$ was $Y(s)$, which is given by

$$Y(s) = \left(\int_{-\infty}^{\infty} x(t)e^{-st} dt \right) \left(\int_{-\infty}^{\infty} h(\lambda)e^{-s\lambda} d\lambda \right)$$

Consider some s_i such that $s_i \in R_X \cap R_H$. It is clear that this product will **definitely converge** for this s_i . So, region of convergence includes this intersection.

$$R_X \cap R_H \subset R_Y$$

The ROC *can* include points other than those lying in this intersection.

7.5.3. New way to characterize the LSI systems

Consider a LSI system with an impulse response $h(t)$. Now, if the input signal and output signal are respectively $x(t)$ and $y(t)$, then we have

$$y(t) = x(t) * h(t)$$

. Hence, by the convolution property of the Laplace transform, we have

$$y(t) \xrightarrow{\mathcal{L}} X(s)H(s)$$

with the Region of convergence $R_Y \subset R_X \cap R_H$. Hence we can write

$$H(s) = \frac{Y(s)}{X(s)}$$

And here $H(s)$ is independent of input. This $H(s)$ together with its ROC R_H characterize the linear shift invariant system completely. If we can invert this Laplace transform, then we know the impulse response, and hence know everything about the system.

For this reason, $H(s)$ is called the *system function* of the system S . So, a system function of an LSI system is equal to the Laplace transform of the impulse response of that system, together with its ROC.

7.5.4. Convolution property for the z-transform

Let $x[n]$ and $h[n]$ be two sequences such that

$$x[n] \xrightarrow{\mathcal{Z}} X(z) ; R_X$$

7. chapter

and

$$h[n] \xrightarrow{z} H(z) ; R_H$$

Let $y[n] = (x * h)[n]$, where (*) implies convolution. Hence,

$$y[n] = \sum_{m=-\infty}^{\infty} x[m] \cdot h[n-m]$$

Therefore,

$$Y(z) = \sum_{n=-\infty}^{\infty} y[n] z^{-n} = \sum_{n=-\infty}^{\infty} \left(\sum_{m=-\infty}^{\infty} x[m] \cdot h[n-m] \right) z^{-n}$$

Replace $(n - m)$ by l . Hence $n = l + m$. Also, for a fixed m , as n goes from $-\infty$ to ∞ , l also goes from $-\infty$ to ∞ . Therefore

$$Y(z) = \sum_{m=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} x[m] \cdot h[l] z^{-(m-l)}$$

$$Y(z) = \sum_{m=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} x[m] \cdot h[l] z^{-(m+l)}$$

$$Y(z) = \left(\sum_{n=-\infty}^{\infty} x[n] z^{-n} \right) \left(\sum_{m=-\infty}^{\infty} h[m] z^{-m} \right)$$

Hence, we have the following property:

$$Y(z) = X(z) \cdot H(z)$$

For $z \in (R_X \cap R_H)$, $Y(z)$ definitely converges. Therefore, when two sequences are convolved, their corresponding z-transforms are multiplied. And the region of convergence of this multiplication is *at least* the intersection of the regions of convergence of the corresponding original signals.

Now, let us consider an example where the ROC is a null set.

ROC is a Null Set

Let us take an example of this.

$$x[n] = 2^n, \forall n$$

$x[n]$ can be simplified as

$$x[n] = 2^n u[n] + 2^n u[-n-1]$$

what we are trying to say is that for n from $-\infty$ to -1 , $x[n] = 2^n u[-n-1]$ for n from 0 to ∞ , $x[n] = 2^n u[n]$ Now

$$2^n u[n] \xrightarrow{z} \frac{1}{1 - 2z^{-1}} |z| > 2$$

and

$$2^n u[-n-1] \xrightarrow{z} \frac{-1}{1 - 2z^{-1}} |z| < 2$$

Adding these two we get

$$\begin{aligned} x[n] &\xrightarrow{z} X(z) \\ X(z) &= 0; \end{aligned}$$

But the region of convergence is $\text{null}(\emptyset)$. This means that $X(z)$ has no region of convergence. So, in the traditional sense $x[n] = 2^n, \forall n$ has no z-transform.

Now, we had said earlier that the ROC of the Laplace (or z-) transform of a convolution is at least a subset of $R_X \cap R_H$. Let us two examples, one in which it is indeed a subset, and other in which it contains more points than the subset.

7.5.5. Examples

ROC is a subset of $R_x \cap R_H$

Let $x(t)$ and $h(t)$ be two signals given by

$$x(t) = e^{2t}u(t)$$

$$h(t) = e^{3t}u(t)$$

Then, we have

$$\begin{aligned} [x * h](t) &= \int_{-\infty}^{\infty} e^{2\tau}u(\tau)e^{3(t-\tau)}u(t-\tau)d\tau \\ &= \int_0^t \{e^{2\tau}e^{3(t-\tau)}d\tau\} \text{ for } t \geq 0 \\ &= u(t) \int_0^t \{e^{2\tau}e^{3t}e^{-3\tau}d\tau\} \\ &= e^{3t}u(t) \int_0^t e^{(2-3)t}d\tau \\ &= e^{3t}u(t) \int_0^t e^{(-\tau)}d\tau = (e^{3t} - e^{2t})u(t) \end{aligned}$$

Here we can see that the *convolution of these two exponential is actually a linear combination of those exponential*. This is a very important property of exponential. Now we can find the *Laplace transform* using linearity.

$$Y(s) = \frac{1}{(3-s)}e^{(3-s)t}\Big|_0^\infty - \frac{1}{(2-s)}e^{(2-s)t}\Big|_0^\infty$$

$$\mathbf{Y}(s) = \frac{1}{(s-3)} - \frac{1}{(s-2)}$$

$$R_Y = \text{Re}(s) > 2 \cap R_e(s) > 3$$

$$\mathbf{R}_Y : \text{Re}(s) > 3$$

We can simplify it further

$$Y(s) = \frac{1}{(s-2)(s-3)}$$

And here the Laplace transform of $y(t)$ is product of the Laplace transforms of the individual exponential.

Expansion of ROC

Let $X(z)$ and $H(z)$ be the z-transforms of the sequences $x[n]$ and $h[n]$ respectively, with regions of convergence R_X and R_H respectively. It may be possible that the function $X(z) \cdot H(z)$ will have a region of convergence which is larger than $R_X \cap R_H$. Let us take an example. Let there be two sequences $x[n]$ and $h[n]$

$$x[n] = 2^n u[n]$$

and

$$h[n] = \{1, -2\}$$

that is,

$$h[n] = \begin{cases} 1 & : n = 0 \\ 2 & : n = 1 \\ 0 & : \text{otherwise} \end{cases}$$

We know that this $x[n]$ has the ROC given by $R_X : |z| > 2$. Now,

$$\begin{aligned} H(z) &= \sum_{n=-\infty}^{\infty} h[n] z^{-n} \\ &= 1 - \frac{2}{z} \end{aligned}$$

This $H(z)$ has the ROC $R_H : |z| \neq 0$.

Hence we have,

$$R_X \cap R_H : |z| > 2$$

Now, let

$$\begin{aligned} y[n] &= (x * h)[n] \\ &= \sum_{m=-\infty}^{\infty} x[m] \cdot h[n-m] \\ &= \sum_{m=-\infty}^{\infty} x[n-m] \cdot h[m] \\ &= x[n]h[0] + x[n-1]h[1] \\ &= 2^n u[n] - 2 \cdot 2^{n-1} u[n-1] \end{aligned}$$

Hence,

$$y[n] = \begin{cases} 1 & : n = 0 \\ 0 & : \text{otherwise} \end{cases}$$

Hence,

$$y[n] = \delta[n]$$

and

$$Y(z) = 1$$

Hence $Y(z)$ has the entire complex plane as the region of convergence, whereas $R_X \cap R_H = |z| > 2$. This is an example of the expansion of ROC (region of convergence) beyond intersection.

7.5.6. System Function

A LSI system for which the impulse response has

- A Laplace transform if on the continuous independent variable, or
- A z-transform if on the discrete independent variable

then it is said to have a *system function*.

The system function for a continuous independent variable system (respectively a discrete independent variable system) is a Laplace transform (respectively a z-transform) and therefore has both, an expression (in s -domain or in z -domain at respective place) and a Region of Convergence (ROC).

Earlier we described the systems by their frequency response, which was a function of the angular frequency (ω) or cycles per second frequency (f). but there was no notion of ROC. The notion of ROC becomes important because we are dealing with systems, which are possibly unstable, which could be non-causal or both.

7.5.7. Importance of Region of Convergence

We have seen cases where the same expression with different ROC will give you different functions underlying and that would be true for impulse response also.

So, the ROC is always there, associated with the system function, whether we state it explicitly or not. In some cases the ROC will be obvious from the context.

Let us take an example to understand all of this.

Consider two systems with impulse responses

$$h_1(t) = e^{2t} \cdot u(t)$$

$$h_2(t) = -e^{2t} \cdot u(-t)$$

,respectively. This implies that and

$$H_1(s) = \frac{1}{s-2} ROC : Re(s) > 2$$

$$H_2(s) = \frac{1}{s-2} ROC : Re(s) < 2$$

So, we see that the system with impulse response $h_1(t)$ is causal and the system with impulse $h_2(t)$ is non-causal. Both have the same expression for the system function but have different ROC. Therefore, if we are told that the system is causal, given that the expression is $\frac{1}{(s-2)}$ then there is only one possible ROC i.e. $Re(s) > 2$. In many cases, the context might make the ROC clear. But the ROC is always associated with the system function. The expression and the ROC together constitute the system function.

7.5.8. System Function as Ratio of transforms

We have seen for an LSI system, the Laplace transforms of the input, output and the impulse response are related by

$$H(s) = \frac{Y(s)}{X(s)}$$

7. chapter

This is also true for discrete systems, owing to the convolution property of the z-transform.

$$H(z) = \frac{Y(z)}{X(s)}$$

Hence,

$$H(s) = \frac{\text{Output Laplace transform}}{\text{Input Laplace transform}} \text{ for continuous variable}$$

and

$$Y(z) = \frac{\text{Output z-transform}}{\text{Input z-transform}} \text{ for discrete variable}$$

provided that the transforms exists.

We cannot talk about a LSI system which has a system function if the input and/or the output do not have a Laplace transform or z-transform. In that case this definition is not meaningful.

In case we have an LSI system with a system function but the input does not have a Laplace transform or z-transform, then to find the output we have to take the inverse of the system function to get the impulse response and then convolve the impulse response with the input to get the output.

7.5.9. A Neat Way to Convolve

The operation of convolution, which can be tedious sometimes, can be made easier using the Laplace or z-transform. Multiply the Laplace transforms (respectively z-transforms) of the signals to be convolved in the continuous (respectively discrete) independent variable. Then take the inverse Laplace transform (respectively inverse z-transform), keeping in mind the ROC.

The Laplace transform and z-transform allows you to convolve conveniently if the inversion of the Laplace transform or z-transform is convenient. It makes sense only when it is easy to go the s -domain or z -domain from the natural domain and if it is easy to come out of the s -domain or z -domain to the natural domain. Otherwise it is better to convolve in the natural domain directly.

7.5.10. Conclusion

In this session we commented about region of convergence(ROC) of the output of the LSI system which is convolution of the input and impulse response of the system. We saw that how Laplace transform of impulse response together with its ROC gives the complete information about the system. We talked about the System Function also. In the next lecture, we will see how to deal with a class of typical impulse responses found in Nature, which are called "poly-ex" terms.

7.6. Module 4: Lecture 6

Dealing with Poly Ex Terms and Partial Fraction Expansion

7.6.1. Introduction

In the previous session you were introduced to the Systems function for LSI systems which is essentially defined as the ratio of the Laplace transform of the output

function to the input function for continuous time variable systems and ratio of Z-transform of the output function to the input function for the discrete time variable system. The impulse response of most of the natural engineering systems is typically a linear combination of “Poly ex” terms. In this session you are going to learn about how to evaluate the Laplace Transform of “Poly ex” functions using the Differentiation property of Laplace transform. We will also discuss about how to evaluate the Laplace inverse of a product of factors using Partial fraction expansion method.

7.6.2. Polynomial exponential function

The “Poly Ex” function is as the name suggests Polynomial multiplied by Exponential function. For example, $(2t^3 + 3t + 5)e^{-4t}u(t)$.

The general form of the Poly ex function can be written as $(\sum_{l=0}^{l=M} a_l t^l) e^{-\alpha t} u(t)$. Here our exponential parameter α can be real or complex i.e. $\alpha = \alpha_R + j\alpha_I$ where α_R the is real part of α which is alright; whereas α_I is the imaginary part of α which would give you the complex part in the expression. If the overall impulse response is real then α must come together with α complex conjugates so in other words if you have one term of this form then you must have another term of the form $(\sum_{l=0}^{l=M} a_l t^l) e^{-\bar{\alpha}t} u(t)$.

Here we have taken right sided function $u(t)$ which is 0 up to $t = 0$ and then equal to unity for all $t > 0$. But you can also take a left sided function i.e. $u(-t)$ which is equal to unity for all $t < 0$ and then 0 for all $t > 0$. In general we may take any of right or left sided Poly ex function and also we can take our zero at any point on t -axis i.e. $u(t - t_0)$ which is equal to 0 up to $t = t_0$ and then equal to 1. You can fix the zero where you like but you can't do it if there are multiple Poly ex terms because you may not have this same point from which they start. So the statement that you can fix your zero where you like is meaningful to the extent that you are dealing with just one term. So you could have right sided or left sided signals, unfortunately you cannot have signals that last all over time; we saw that, if you had e^2t for all t it has no Laplace transform, in the general sense of Laplace transforms using functions you are not using impulses and things like that. If one wants to deal with functions in the s plane or in the Z plane then exponentials lasting all over time have no Laplace transform or Z-transform. So you need either left sided or right sided signals.

7.6.3. Differentiation property for Laplace transform

You know how to invert the Laplace transforms of the form $\frac{1}{(s+\alpha)}$. These kind of things we have been doing extensively. What would be the Laplace inverse if we multiply this kind of functions $\frac{1}{(s+\alpha)}$ by a polynomial? For a polynomial of degree 0 (a constant) there is no problem; but what if we multiply it by t or t^2 ? For that we need to establish our next property called property of differentiation for Laplace transform.

Let $x(t)$ has the Laplace transform $X(s)$ with a Region of Convergence(ROC) R_x . The question is what would be the Laplace inverse if we differentiate $X(s)$ in the s domain with almost the same ROC R_x except some isolated contours. Let's write

7. chapter

down $X(s)$ in terms of $x(t)$ and then differentiate both sides with respect to s ,

$$X(s) = \int_{-\infty}^{\infty} x(t)e^{-st} dt, \text{ ROC : } R_x$$

$$\frac{dX(s)}{ds} = \int_{-\infty}^{\infty} (-t)x(t)e^{-st} dt \text{ ROC : } R_x$$

Here, we are assuming that $X(s)$ is differentiable with respect to s and we can take the derivative inside the integral on the integrand which is a function of s . Now, multiply both sides by -1 and see now we have the Laplace transform of $t.x(t)$ which is just negative of the derivative of Laplace transform of $x(t)$.

$$-\frac{dX(s)}{ds} = \int_{-\infty}^{\infty} (t.x(t))e^{-st} dt \text{ ROC : } R_x$$

For example, let's find the Laplace transform of $te^{-3t}u(t)$.

You know the Laplace transform of $e^{-3t}u(t)$ which is $\frac{1}{(s+3)}$. So by using the property of differentiation we just derived, it should be $-\frac{d(\frac{1}{(s+3)})}{ds} = \frac{1}{(s+3)^2}$. Now we are quite equipped to deal with the Laplace transform of Poly ex terms. Let's put down a procedure. Let's take the example which we took earlier.

$$x(t) = \left(\sum_{l=0}^{l=M} a_l t^l \right) e^{-\alpha t} u(t)$$

Laplace transform of this is equal to the sum of individual Laplace transforms of each term which can be calculated using the differentiation property repeatedly. Let's take Laplace transform of a general term $a_l t^l e^{-\alpha t} u(t)$. Laplace transform of $e^{-\alpha t} u(t)$ is $\frac{1}{(s+\alpha)}$ (take it as $F(s)$). After multiplying by t it should be $-\frac{dF(s)}{ds}$ and so after multiplied by t repeatedly for $l - 1$ times it would be $(-1)^l \frac{d^l F(s)}{ds^l}$. Also the multiplication by a constant is not a big deal since Laplace transform follows the homogeneity principle. So the final expression for the Laplace transform of a Poly ex function $x(t)$ can be written as follows:

$$X(s) = \sum_{l=0}^{M} \frac{a_l l!}{(s + \alpha)^{l+1}}$$

7.6.4. Region of Convergence

Using the differentiation property we are able to deal with Laplace transforms of the form $\frac{1}{(s+\alpha)^M}$ (where M is a positive integer) with $Re(s + \alpha) > 0$ or $Re(s + \alpha) < 0$. We have two different possible regions of convergence for $Re(s + \alpha) > 0$ and $Re(s + \alpha) < 0$. There is a very simple principle to associate these with the so called right sided and left sided signals. In the complex plane there are two explicit regions comprising $Re(s) > Re(-\alpha)$ and $Re(s) < Re(-\alpha)$; both are separated by a critical line $Re(s) = Re(-\alpha)$. You know that for existence of the Laplace transform, the input Poly ex signal should be either right sided or left sided. Here the rule is simple; the right side region of convergence corresponds to the right sided signal and the left side region of convergence corresponds to the left sided input

signal i.e. if the input signal is something multiplied by $u(t)$ then ROC should be $\text{Re}(s) > \text{Re}(-\alpha)$. Also for right or left sided signal the shifting in time does not affect the ROC.

Now, what happens when Laplace transform has more than one such term? Let's look at that possibility now. Suppose Laplace transform is of the following form $\frac{1}{(s+\alpha)(s+\beta)}$, obviously with $\alpha \neq \beta$. Here there are three possible regions of convergence : $\text{Re}(s) < \text{Re}(-\alpha)$, $\text{Re}(-\alpha) < \text{Re}(s) < \text{Re}(-\beta)$ and $\text{Re}(s) > \text{Re}(-\beta)$ (assuming $\text{Re}(-\alpha) < \text{Re}(-\beta)$); which are separated by two critical vertical lines $\text{Re}(s) = \text{Re}(-\alpha)$ and $\text{Re}(s) = \text{Re}(-\beta)$. It is very clear that taking any one region of convergence it falls into one of those two possible regions of convergence with respect to a specific vertical line. You can break this Laplace transform into terms that contains only one of these vertical lines by using Partial fraction expansion method.

7.6.5. Partial Fraction Expansion

The Laplace transform $X(s) = \frac{1}{(s+\alpha)(s+\beta)}$ is expected to be of the form $\frac{A}{(s+\alpha)} + \frac{B}{(s+\beta)}$. It's very simple to determine the constants A and B here. Let

$$X(s) = \frac{1}{(s+\alpha)(s+\beta)} = \frac{A}{(s+\alpha)} + \frac{B}{(s+\beta)}$$

Now, multiply both the sides by $(s+\alpha)$, then it is

$$\frac{1}{(s+\beta)} = A + \frac{B(s+\alpha)}{(s+\beta)}$$

Now, put $s = -\alpha$, so we get $A = \frac{1}{(\beta-\alpha)}$. Similarly you can find $B = \frac{1}{(\alpha-\beta)}$.

So, you when we have simple factors like this there is no problem; Partial fraction is simple. But what do we do when we have multiple factors? So, let's see that case too. You could have $X(s) = \frac{1}{(s+\alpha)^2(s+\beta)}$. In this case there are several ways to decompose this into partial fractions, but here the type we wish to use is of the form

$$X(s) = \frac{1}{(s+\alpha)^2(s+\beta)} = \frac{A}{(s+\alpha)} + \frac{B}{(s+\beta)} + \frac{C}{(s+\alpha)^2}$$

Here, to find B and C is not a big deal; to find B multiply both the sides by $s+\beta$ and then put $s = -\beta$ in the equation. Similarly, to find C multiply both sides by $(s+\alpha)^2$ and then put $s = -\alpha$ in the equation. But to find A we have to do something different.

After multiplying both sides by $(s+\alpha)^2$,

$$\frac{1}{(s+\beta)} = A(s+\alpha) + \frac{B(s+\alpha)^2}{(s+\beta)} + C$$

Now, we need to differentiate both sides with respect to s and then put $s = -\alpha$. So the terms containing B and C will vanish and only the term containing A would survive. After doing so you will get $A = -\frac{1}{(\beta-\alpha)^2}$ and of course, B and C will be respectively $\frac{1}{(\alpha-\beta)^2}$ and $\frac{1}{(\beta-\alpha)}$.

7.6.6. Conclusion

So now we have well equipped to study the Laplace transform of a “poly ex” term, so at least we are now competent to deal with a large number of engineering systems which have such impulse responses. All the typical electrical networks or mechanical systems using masses and frictional parts and springs would typically have these kinds of responses. So it’s very clear that we now know how to recognize what happens when you have a square or a cube.

We can find the Laplace inverse of a product of factors easily after decomposing it into the terms of the form $\frac{A}{(s+\alpha)^n}$ which is the Laplace transform of a Poly ex function i.e. $\frac{At^n e^{-\alpha t}}{n!} u(t)$.

7.7. Module 4: Lecture 7

Dual of the Differentiation Property and Rational Laplace Transforms

7.7.1. Introduction

In the last lecture, we defined a specific class of Laplace transforms; the rational Laplace transforms. In this lecture, we will look in to rational Laplace transforms in greater depth. We will look at the dual of the differentiation property, by giving a formal expression for the inverse Laplace transform. Finally, we will emphasize the importance of rational Laplace transforms in the context of Linear Shift Invariant systems.

7.7.2. Rational Laplace Transforms

A rational Laplace transform is the one which can be expressed as a ratio of two *finite length* series in the Laplace variable s , which means a ratio of a numerator finite length series in s , to a denominator finite length series in s . For example, consider the following Laplace transform:

$$X(s) = \frac{1}{(s+2)(s+3)} = \frac{1}{s^2 + 5s + 6}$$

Note that we can have both positive and negative powers of s in the series.

Now, as discussed in the previous lecture, we can simplify these transforms further by the use of partial fractions. So, in this particular case, we have

$$\frac{1}{(s+2)(s+3)} = \frac{A}{(s+2)} + \frac{B}{(s+3)}$$

where A (respectively B) can be evaluated by multiplying this equation by $(s+2)$ (respectively $(s+3)$) and putting $s = -2$ (respectively $s = -3$). It is easily verified that $A = 1$ and $B = -1$. Hence,

$$\frac{1}{(s+2)(s+3)} = \frac{1}{(s+2)} - \frac{1}{(s+3)}$$

Now, this is only the expression of the Laplace transform. To completely specify the Laplace transform, we must also specify the ROC. Note that there are three possibilities for the ROC, which is specified in the table below.

	Region of Convergence	Inverse Laplace Transform of $\frac{1}{(s+2)} - \frac{1}{(s+3)}$
First Region	$Re(s) < -2 \text{ & } Re(s) < -3$ $\Rightarrow Re(s) < -3$	Take both signals left sided. Hence $x(t) = (-e^{-2t} + e^{-3t})u(-t)$
Second Region	$Re(s) < -2 \text{ & } Re(s) > -3$ $\Rightarrow -3 < Re(s) < -2$	$x(t) = -e^{-2t}u(-t) - e^{-3t}u(t)$
Third Region	$Re(s) > -2 \text{ & } Re(s) > -3$ $\Rightarrow Re(s) > -2$	$x(t) = e^{-2t}u(t) - e^{-3t}u(t)$

As we know, when both $Re(s) < -2$ and $Re(s) < -3$, both the corresponding signals will be left sided. Hence we will get $u(-t)$ in both the cases. For the second region, as $Re(s) > -3$, the signal corresponding to $1/(s+3)$ will be right sided while the one corresponding to $1/(s+2)$ will be left sided, and similarly for the third region.

Now, as discussed in the earlier lecture, we can deal with terms involving higher orders of $(s+\alpha)$ in the denominator, by expanding in partial fractions accordingly and evaluate the coefficients of the various powers of $(s+\alpha)$ by successive differentiation.

Now, let's see why it is important to study rational Laplace transforms. For that, we have to study one more property of the Laplace transform, which, in some sense, is the dual of the differentiation property which was discussed earlier.

7.7.3. Inverse of the Laplace Transform

We had seen what is the signal in time domain corresponding to the derivative of the Laplace transform. The question asked was as follows. If

$$x(t) \xrightarrow{\mathcal{L}} X(s)$$

Then what is $y(t)$ such that

$$y(t) \xrightarrow{\mathcal{L}} \frac{dX(s)}{ds}$$

It was proved that $y(t) = -t \cdot x(t)$. Hence, differentiating in the Laplace domain results in the multiplication by the independent variable (with a minus sign) in the time domain. Similarly, we can ask the question, what will be the Laplace transform of $\frac{dx(t)}{dt}$?

To answer this question, it would be convenient if we could express $x(t)$ in terms of $X(s)$, that is, provide a formal expression for the inverse of the Laplace transform.

Formal Inversion of the Laplace Transform

Let us have a look at how we arrived at the Laplace transform in the first place. We first multiplied the signal by a suitable decaying exponential, so that the signal

7. chapter

becomes absolutely integrable, and then took the Fourier transform. So

$$x(t) \xrightarrow{\text{multiplication by } e^{\sigma t}} x(t).e^{\sigma t} \xrightarrow{\mathcal{F}} X(s)$$

where \mathcal{F} denotes the Fourier transform.

Hence,

$$x(t) \xrightarrow{\mathcal{L}} \int_{-\infty}^{+\infty} x(t)e^{\sigma t}e^{-j\Omega t} dt$$

Hence, we can write

$$X(s) = \int_{-\infty}^{+\infty} x(t)e^{-st} dt \text{ with } s = -\sigma + j\Omega$$

Now, this itself tells us how to proceed to evaluate $x(t)$ from $X(s)$. Essentially, we have to go backwards. Consider $X(s)$ with $s = -\sigma + j\Omega$ with fixed value of σ , lying in the region of convergence. If we take its inverse Fourier transform, we should get back $x(t).e^{\sigma t}$. Hence,

$$X(-\sigma + j\Omega) \xrightarrow{\mathcal{F}^{-1}} x(t).e^{\sigma t}$$

Now, multiplying this output by $e^{-\sigma t}$ should give $x(t)$. Now, writing the expression for the inverse Fourier transform, we get,

$$x(t) = \left\{ \frac{1}{2\pi} \int_{-\infty}^{+\infty} X(-\sigma + j\Omega)e^{j\Omega t} d\Omega \right\} .e^{-\sigma t}$$

Now, since the integral is over Ω , we can take the $e^{-\sigma t}$ inside the integral. So,

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} X(-\sigma + j\Omega)e^{(-\sigma+j\Omega)t} d\Omega$$

for a fixed $-\sigma$ lying in the region of convergence. Now, $s = -\sigma + j\Omega$. Also, as σ is fixed, we have $ds = j.d\Omega$. As Ω goes from $-\infty$ to $+\infty$, s goes from $-\sigma - j\infty$ to $-\sigma + j\infty$. And hence, we have our formal expression for the inverse Laplace transform:

$$x(t) = \frac{1}{2\pi j} \int_{-\sigma-j\infty}^{-\sigma+j\infty} X(s)e^{st} ds$$

This integral is over a contour in the complex plane in which s lies. This contour is essentially a vertical line corresponding to $Re(s) = -\sigma$. Now, to simplify the notation, we will write the limits of integration as just $-\infty$ to $+\infty$, keeping in mind that the integration over the complex variable s is on the vertical contour $Re(s) = -\sigma$.

Hence,

$$x(t) = \frac{1}{2\pi j} \int_{-\infty}^{+\infty} X(s)e^{st} ds$$

Now that we have obtained a formula for the inversion of the Laplace transform, we can see what happens when we differentiate $x(t)$.

Interestingly, the formal inversion is useful *not* for inverting the Laplace transforms, in general. As we will see, in most of the cases, the inverse Laplace transform is obtained strategically by using experience. But, this formal inversion of Laplace transform can provide us a useful insight, as shown in the next subsection.

Dual of the Differentiation Property

We had seen that if

$$x(t) \xrightarrow{\mathcal{L}} X(s)$$

then

$$-t \cdot x(t) \xrightarrow{\mathcal{L}} \frac{dX(s)}{ds}$$

We would like to see what happens then the signal is differentiated in the time domain instead. In short, what is the Laplace transform of $\frac{dx(t)}{dt}$?

We can proceed with the inversion formula obtained at the end of the previous section.

$$x(t) = \frac{1}{2\pi j} \int_{-\infty}^{+\infty} X(s) e^{st} ds$$

Where the integration is on a vertical line in the ROC. Let us denote the ROC by L . We can now differentiate with respect to t on both sides to obtain

$$\frac{dx(t)}{dt} = \frac{1}{2\pi j} \frac{d}{dt} \int_{-\infty}^{+\infty} X(s) e^{st} ds = \frac{1}{2\pi j} \int_{-\infty}^{+\infty} X(s) \frac{d(e^{st})}{dt} ds = \frac{1}{2\pi j} \int_{-\infty}^{+\infty} s X(s) e^{st} ds$$

Hence, we can see that

$$\frac{dx(t)}{dt} = \frac{1}{2\pi j} \int_{-\infty}^{+\infty} \{s X(s)\} e^{st} ds$$

Now, hoping that the integral would converge on the same ROC L , we can see that it is essentially the inverse Laplace transform of $s \cdot X(s)$. So,

$$\frac{dx(t)}{dt} \xrightarrow{\mathcal{L}} s X(s)$$

Multiplication by s is normally not going to change the region of convergence too much, except for possibly the extremities.

Hence, while differentiating in the Laplace domain results in the multiplication of the independent variable (with a minus sign) in the time domain, the differentiation in the time domain results in the multiplication of the independent variable in the Laplace domain. Hence, this is a dual property.

Importance of the Rational Laplace Transform

The dual of the differentiation property seen above gives us a new insight in to why the rational Laplace transforms are important. Every time you take a derivative in time you're multiplying by s . The other way to look at it is that every time you're multiplying by s , you're taking the derivative in time. That is, $sX(s)$ will be the Laplace transform of $\frac{dx(t)}{dt}$, and $s^2 X(s)$ will be the Laplace transform of $\frac{d^2 x(t)}{dt^2}$, and so on. Let us see the importance of this, by looking at the following example.

Consider a LTI system, with an impulse response $h(t)$, having a Laplace transform. Thus, the system function is $H(s)$. So, if the input is $x(t)$ and $y(t)$, assuming both have a Laplace transform, we have,

$$Y(s) = X(s)H(s)$$

7. chapter

or

$$H(s) = \frac{Y(s)}{X(s)}$$

Now, consider the case when the system function is *rational*. Let's say

$$H(s) = \frac{(s+3)}{(s+2)(s+4)}$$

Hence,

$$\begin{aligned} \frac{(s+3)}{(s+2)(s+4)} &= \frac{Y(s)}{X(s)} \\ \implies (s+3)X(s) &= (s+2)(s+4)Y(s) = (s^2 + 6s + 8)Y(s) \end{aligned}$$

Now, all the expressions are of the form of Laplace transform or some integral power of s times the Laplace transform. Hence, using the property derived in the last subsection, we can take the inverse Laplace transform of the equation as follows:

$$\frac{dx(t)}{dt} + 3x(t) = \frac{d^2y(t)}{dt^2} + 6\frac{dy(t)}{dt} + 8y(t)$$

This is a linear constant coefficient differential equation. At the end of the first module in signals and systems, we discussed that linear shift-invariant systems are most typical in natural engineering systems. Hence, we see that a for a system whose impulse response has a rational Laplace transform corresponds to a Linear Constant Coefficient Differential Equation (LCCDE) in the input and output in time domain, which occurs in most engineering systems. Here, we define a *rational system* as the one which is linear and shift invariant, and has an impulse response which has a rational Laplace transform. Hence as rational systems occur most commonly, it is important to study rational Laplace transforms.

There is one more detail which needs to be clarified. We have

$$x(t) \xrightarrow{\mathcal{L}} X(s)$$

$$\frac{dx(t)}{dt} \xrightarrow{\mathcal{L}} sX(s)$$

What happens when there is a negative power of s . That is to say, which signal corresponds to the Laplace transform $\frac{X(s)}{s}$?

To obtain this, we must go backwards in the derivation of the dual property. Hence if we denote $\frac{dx(t)}{dt} = g(t)$ and $sX(s) = G(s)$, then we have

$$g(t) \xrightarrow{\mathcal{L}} G(s)$$

$$\int_{-\infty}^t g(t') dt' \xrightarrow{\mathcal{L}} X(s) = \frac{G(s)}{s}$$

Hence, in dividing by s , we essentially have the inverse of the derivative operator, which is called the running integral.

7.7.4. Irrational Systems

We said in the definition of rational systems, that it should be of the form numerator *finite* series in s upon a denominator *finite* series in s . We have emphasised the condition of the finiteness of the series. Let us take an example of a simple irrational system whose system function can be expanded in a simple infinite series.

Consider a simple delay, whose system description is given by

$$y(t) = x(t - \tau)$$

where τ is a constant. Now, we can easily calculate the impulse response, by putting $x(t) = \delta(t)$. Hence,

$$h(t) = \delta(t - \tau)$$

The system function is thus

$$H(s) = \int_{-\infty}^{+\infty} \delta(t - \tau) e^{-st} dt = e^{-s\tau}$$

This is an irrational system function, as it cannot be expressed as a division of two *finite* series in s .

The problem with irrational systems is that only rational systems are truly realisable, meaning that only rational systems can be realised or implemented with *finite* resources. Of course, an irrational system can be in principle realised using infinite resources, but that's impractical. In practice, wherever irrational systems are needed, they are realised with a certain degree of approximation. For example, the ideal delay system as described above can be approximated by a system which behaves like a constant delay for a large range of inputs, and so on. The approximation can be improved by putting more and more resources, but to achieve the exact system, infinite resources are required.

Even if we cannot write a rational form of $e^{-s\tau}$, note that we can easily expand it in an infinite series, using Taylor's expansion. We have

$$e^{-s\tau} = \sum_{n=0}^{n=\infty} \frac{(-s\tau)^n}{n!}$$

where $n!$ is defined by the following equations:

$$0! = 1$$

$$1! = 1$$

$$n! = n \times (n - 1)!$$

Hence, we have

$$2! = 2 \times 1$$

$$3! = 3 \times 2 \times 1$$

and so on. Now, this series for $e^{-s\tau}$ is infinite and *cannot be converted* to a ratio of finite series. This point is stressed because there can exist infinite series which *can* be expressed as a ratio of finite series. Consider, for example, the series expansion of $1/(1 - s)$.

$$\frac{1}{(1 - s)} = 1 + s + s^2 \dots \infty$$

in the region given by $|s| < 1$. But the series expansion of $e^{-s\tau}$ cannot be expressed in such a form. Hence, the system is irrational.

7.7.5. Conclusion

We have seen why the rational Laplace transforms are important, through the study of the dual of the differentiation property of the Laplace transform. In the coming lectures, we will see more properties of the rational Laplace transforms, which will help us invert them, and also provide new insights into the system.

7.8. Module 4: Lecture 8

System Analysis Using Laplace Transform

7.8.1. Introduction

In the last lecture, we studied regarding rational and irrational systems. We had also seen that rationality of a system was an important property. A general Rational Laplace Transform is of the form $\frac{p}{q}$ where p and q are polynomials in s .

Example : $\frac{s^{-1}+2s+1}{s^{-2}+s^{-1}+3+2s}$

This can be rewritten as $\frac{s^2+2\cdot s^4+s^3}{s+s^2+3\cdot s^3+2\cdot s^4}$.

7.8.2. Poles and Zeroes

Let us denote the polynomial in numerator as $N(s)$ and the polynomial in denominator as $D(s)$. So in the above example $N(s) = s^2 + 2 \cdot s^4 + s^3$ and $D(s) = s + s^2 + 3 \cdot s^3 + 2 \cdot s^4$

Consider the roots of $D(s)$ i.e. the set of values of s for which $D(s) = 0$. In the above case we have,

$$s + s^2 + 3 \cdot s^3 + 2 \cdot s^4 = 0$$

$$s(1 + s + 3 \cdot s^2 + 2 \cdot s^3) = 0$$

which implies either $s = 0$ or the cubic equation in s is zero. We know that we can solve the cubic equation for its roots.

Lets see what happens when $D(s)$ tends to 0. Consider the magnitude of Laplace Transform in s-plane. As $D(s) \rightarrow 0$, the magnitude of Laplace transform tends to infinity which can be signified as a tent rising to infinity in s -plane at that point. The values of s for which $D(s)$ goes to 0 are called as *Poles*.

Now consider $N(s)$ tending to zero. This makes the magnitude of Laplace transform to go to zero which can be signified as a tent which is touching the ground. The values of s for which $N(s)$ goes to 0 are called as *Zeros*.

Its the poles and zeros together which make the rational Laplace Transform.

Example: A rational Laplace transform has a zero at $s = -1$ and two poles at $s = -1$ and $s = -5$. Then the Laplace transform would be of the form $\frac{K(s+1)}{(s+3)(s+5)}$ where K is a constant about which we have no idea.

Lets calculate the inverse Laplace Transform with the help of poles and zeros.

1. Identify the distinct poles.
2. The Laplace Transform is of the form $\frac{N(s)}{D(s)}$ where degree of numerator polynomial is less than the degree of the denominator polynomial.
3. Each distinct pole will give a set of terms of the form $\frac{K}{(s+\alpha)^l}$. Suppose a pole is at $s = -2$ and it occurs 3 times then the terms

$$\frac{a}{(s+2)} + \frac{b}{(s+2)^2} + \frac{c}{(s+2)^3} \quad (7.1)$$

where a, b, c are constants. Now each of these terms can be inverted as per their region of convergence.

4. Suppose the degree of numerator polynomial is greater than the degree of denominator polynomial, then we need to divide $N(s)$ by $D(s)$ using long division method wherein we get a quotient $Q(s)$ and a remainder $R(s)$. So we can write,

$$\frac{N(s)}{D(s)} = Q(s) + \frac{R(s)}{D(s)} \quad (7.2)$$

where degree of $R(s)$ is less than degree of $D(s)$.

Example:

$$\frac{s^2 + s + 1}{s + 1} = s + \frac{1}{s + 1} \quad (7.3)$$

7.8.3. Singularity Functions

In the above expression, we know to calculate the inverse Laplace Transform of $\frac{1}{s+1}$, but we have a problem when we have to find the inverse Laplace Transform of s .

Lets find the inverse Laplace Transform of a constant c .

$$c = \int_{-\infty}^{\infty} x(t)e^{-st} dt$$

which gives,

$$x(t) = c\delta(t)$$

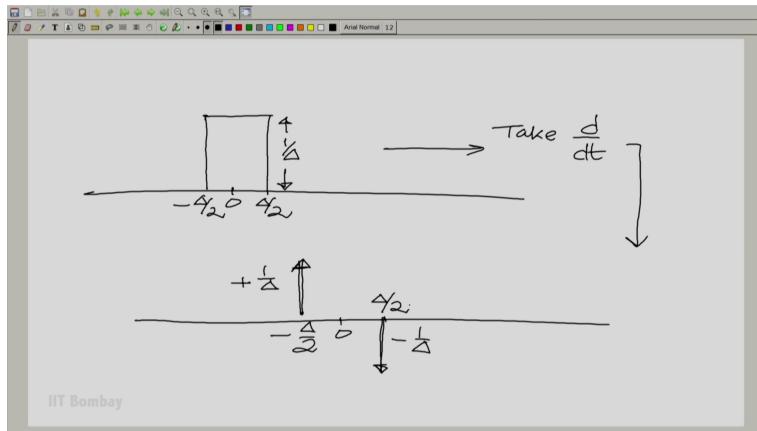
Therefore, inverse Laplace transform of a constant c is an impulse at $t = 0$ of strength c . To find the Laplace transform of $c \cdot s$, we need to differentiate the impulse $\delta(t)$.

$$c\delta(t) \xrightarrow{\mathcal{L}} c ;$$

7. chapter

$$? \xrightarrow{\mathcal{L}} cs ;$$

By using property of Laplace Transform, we know that $? = \frac{d}{dt}(c\delta(t))$. To find the derivative of an impulse, lets look at the impulse carefully.



In the above figure, the top one shows the impulse as Δ tends to zero and the bottom one is the derivative of the impulse. Now as $\Delta \rightarrow 0$, the impulse strength at $-\frac{\Delta}{2} \rightarrow \infty$ and $\frac{\Delta}{2} \rightarrow -\infty$ and these two come closer to each other. This is what we call as *Doublet*. Suppose we let this doublet act upon a differentiable function $x(t)$ at point $t = t_0$.

$$\int_{-\infty}^{\infty} \frac{x(t)}{\Delta} (\delta(t - (t_0 - \frac{\Delta}{2})) - (\delta(t - (t_0 + \frac{\Delta}{2})))) dt$$

which implies

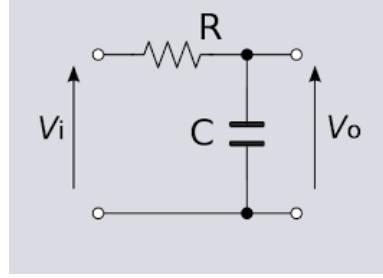
$$\frac{x(t_0 - \frac{\Delta}{2}) - x(t_0 + \frac{\Delta}{2})}{\Delta}$$

As Δ tends to zero, the above expression will become the negative of the derivative of $x(t)$ at $t = t_0$. So we found by applying doublet over a differentiable function, we got the derivative of the function at that point. This doublet is usually called as *Unit Doublet* as it is trying to be unit impulse as $\Delta \rightarrow 0$.

We can extend this idea to find the inverse Laplace Transform of s^2 . Here we need to do double differentiation of impulse, which implies differentiating the doublet. These generalised functions are called *Singularity Functions*. It is easy to understand the effect of these impulses over functions rather than studying them in detail.

7.8.4. RC Circuit

Now lets deal with the example of RC circuit. Let us find the unit step response using its Laplace Transform. Consider a RC circuit as shown in the below figure.



Let the voltage across the resistor R be $V_R(t)$ and the current passing through it be $i_R(t)$. We have

$$V_R(t) = i_R(t)R$$

Taking Laplace Transform on both sides we get,

$$V_R(s) = i_R(s)R$$

$$\frac{V_R(s)}{i_R(s)} = R$$

where R is known as Laplace domain Impedance of Resistance.

Now consider the capacitor in the RC circuit. Let the voltage across the capacitor of capacitance C be $V_c(t)$ and the current passing through it be $i_c(t)$. Then we have,

$$i_c(t) = C \frac{dV_c(t)}{dt}$$

Taking Laplace Transform on both sides we get,

$$I_c(s) = CsV_c(s)$$

$$\frac{V_c(s)}{I_c(s)} = \frac{1}{Cs}$$

where $\frac{1}{Cs}$ is known as Laplace domain Impedance of Capacitance.

Laplace domain Impedance with Laplace Transform will help to handle wide class of circuits consisting of resistances, inductors, capacitors etc.

Consider the RC circuit. Let $V_i(t)$ be input voltage and $V_o(t)$ be the output voltage across capacitor. Then we have,

$$V_i(t) = RC \frac{dV_o(t)}{dt} + V_o(t)$$

Taking Laplace Transform on both sides we get,

$$V_i(s) = RCsV_o(s) + V_o(s)$$

$$\frac{V_o(s)}{V_i(s)} = \frac{1}{1 + RCs}$$

7. chapter

Considering $RC = \tau$ (Time Constant),

$$\frac{V_o(s)}{V_i(s)} = \frac{1}{1 + s\tau}$$

where $\frac{1}{1+s\tau}$ is known as system function. Now this system function has one pole at $s = -\frac{1}{\tau}$

Since we have a single pole, there are two possible Region of Convergences(ROC). One is Right sided impulse response which has a region of convergence from $(-\frac{1}{\tau}, \infty)$. The other is the Left sided impulse response which has a region of convergence from $(-\infty, -\frac{1}{\tau})$. Considering the system to be causal, we need to consider only the Right sided impulse response which has a region of convergence from $(-\frac{1}{\tau}, \infty)$.

In the RC circuit for the unit step input case, the Laplace Transform of unit step input is given by

$$u(t) \xrightarrow{\mathcal{L}} \int_{-\infty}^{\infty} u(t)e^{-st}dt = \int_0^{\infty} e^{-st}dt = \frac{1}{s}; \quad R_x = Re(s) > 0$$

Therefore we have got $V_i(s) = \frac{1}{s}$. Substituting this in the system function mentioned above, we get

$$V_o(s) = \frac{1}{s(1 + s\tau)}$$

Expanding the above expression using partial fractions, we get

$$V_o(s) = \frac{1}{s} - \frac{1}{s + \frac{1}{\tau}}$$

Here ROC will be the intersection of region of convergence of each terms. The first in the RHS of the above expression has a ROC of $Re(s) > 0$ while the other term has a ROC of $Re(s) > -\frac{1}{\tau}$. So taking intersection of these two ROC's, we get the final ROC which comes out to be $Re(s) > 0$.

Taking Inverse Laplace Transform on both sides we get,

$$V_o(t) = u(t) - e^{-\left(\frac{t}{\tau}\right)}u(t) = u(t)\left(1 - e^{-\left(\frac{t}{\tau}\right)}\right)$$

Clearly we can see that the above expression is indeed the step response of RC circuit which we had got earlier. Going ahead we can now calculate the response of a ramp input (a linearly increasing input) given by $V_i(t) = tu(t)$. The Laplace transform of $tu(t)$ comes out to be $\frac{1}{s^2}$. Now to calculate the ramp response we can follow two methods:

1. One is the above mentioned method i.e. calculate the partial fractions and perform inverse Laplace Transform on them.
2. The other is done by cascading the systems using running integrator. First a unit step input is given to RC Circuit and the output of RC circuit is given to the input of the running integrator.

We can verify that in the both approaches we get the same answer.

7.9. Module 4: Lecture 9

Region of Convergence(ROC), Convolution Property and System Function

7.9.1. Differentiation property of *z-transform*

Suppose we have a sequence $x[n]$ and

$$x[n] \xrightarrow{z} X(z); \text{ ROC : } R_X$$

$$X(z) = \sum_{n=-\infty}^{\infty} x[n] \cdot z^{-n}$$

Let us consider $\frac{d}{dz} X(z)$ with the same ROC, except for some possible boundaries.

$$X(z) = \sum_{n=-\infty}^{\infty} x[n] \cdot z^{-n}$$

Here we will differentiate term by term, and if the derivative converges

$$\frac{d}{dz} X(z) = \sum_{n=-\infty}^{\infty} x[n] \cdot (-n) \cdot z^{-n-1}$$

Now, multiply both sides by $(-z)$

$$(-z) \cdot \frac{d}{dz} X(z) = \sum_{n=-\infty}^{\infty} x[n] \cdot (n) \cdot z^{-n}$$

The R.H.S. is essentially the *z-transform* of ' $n \cdot x[n]$ '. Therefore,

$$x[n] \xrightarrow{z} X(z); \text{ ROC : } R_X$$

$$n \cdot x[n] \xrightarrow{z} (-z) \frac{d}{dz} X(z);$$

ROC : essentially the same ROC as $X(z)$ with some boundaries

Let us take an example

$$\left(\frac{1}{2}\right)^n \cdot u[n] \xrightarrow{z} \frac{1}{1 - \frac{1}{2}z^{-1}}; |z| > \frac{1}{2}$$

$$n \cdot \left(\frac{1}{2}\right)^n \cdot u[n] \xrightarrow{z} (-z) \cdot \frac{d}{dz} \left(\frac{1}{1 - \frac{1}{2}z^{-1}}\right); |z| > \frac{1}{2}$$

$$(-z) \cdot \frac{d}{dz} \left(\frac{1}{1 - \frac{1}{2}z^{-1}}\right) = \frac{(-z) \cdot (-1) \cdot (-\frac{1}{2})(z^{-2})}{(1 - \frac{1}{2}z^{-1})^2} = \frac{\frac{1}{2}z^{-1}}{(1 - \frac{1}{2}z^{-1})^2}$$

$$\Rightarrow n \cdot \left(\frac{1}{2}\right)^n \cdot u[n] \xrightarrow{z} \frac{\frac{1}{2}z^{-1}}{(1 - \frac{1}{2}z^{-1})^2}; \text{ ROC : } |z| > \frac{1}{2}$$

7. chapter

Here, we have a square of the denominator with some numerator.

This gives the methodology of dealing with repeated factors in the denominator in the *z-transform*. Its principles are the same as that in the *Laplace transform*.

Let us see the effect of convolution in *z-domain*

$$\begin{aligned} x[n] &\xrightarrow{z} X(z); \text{ ROC : } R_X \\ h[n] &\xrightarrow{z} H(z); \text{ ROC : } R_H \\ (x * h)[n] &\xrightarrow{z} X(z) \cdot H(z); \text{ ROC : atleast } R_X \cap R_H \end{aligned}$$

7.9.2. How to handle rational *z-transforms*?

Now, with the convolution property and the differentiation property of the *z-transforms*, we can solve all rational *z-transforms*.

Let us take an example.

Let $|\alpha| < |\beta|$ and

$$Y(z) = \frac{1}{(1 - \alpha z^{-1})(1 - \beta z^{-1})}; \text{ ROC : } |z| > \max(|\alpha|, |\beta|) \text{ or } |z| > |\beta|$$

Now this can be solved by two methods.

Method 1:

First, decompose $Y(z)$ into partial fractions.

$$\frac{1}{(1 - \alpha z^{-1})(1 - \beta z^{-1})} = \frac{A}{(1 - \alpha z^{-1})} + \frac{B}{(1 - \beta z^{-1})}$$

To find ' A' ', multiply both side by $(1 - \alpha z^{-1})(1 - \beta z^{-1})$ and put $z = \alpha$

$$\Rightarrow A = \frac{1}{1 - \beta \alpha^{-1}}$$

Similarly, to find ' B' ', multiply both side by $(1 - \alpha z^{-1})(1 - \beta z^{-1})$ and put $z = \beta$

$$\Rightarrow B = \frac{1}{1 - \alpha \beta^{-1}}$$

Therefore,

$$Y(z) = \frac{\frac{1}{1-\beta\alpha^{-1}}}{1 - \alpha z^{-1}} + \frac{\frac{1}{1-\alpha\beta^{-1}}}{1 - \beta z^{-1}}$$

$$Y(z) = \frac{\frac{\alpha}{\alpha-\beta}}{1 - \alpha z^{-1}} - \frac{\frac{\beta}{\alpha-\beta}}{1 - \beta z^{-1}}$$

$$Y(z) = \frac{1}{(\alpha - \beta)} \cdot \left(\frac{\alpha}{1 - \alpha z^{-1}} - \frac{\beta}{1 - \beta z^{-1}} \right); \text{ ROC : } |z| > |\beta| \text{ which also implies } |z| > |\alpha|$$

Also 'A' and 'B' can be found by multiplying both sides by $(1 - \alpha z^{-1})(1 - \beta z^{-1})$, which gives

$$\begin{aligned} A \cdot (1 - \beta z^{-1}) + B \cdot (1 - \alpha z^{-1}) &= 1 \\ (A + B) - (A\beta + B\alpha) &= 1 \\ \Rightarrow A + B &= 1 \text{ and} \\ A\beta + B\alpha &= 0 \end{aligned}$$

Solving both the equations we get,

$$A = \frac{1}{1 - \beta\alpha^{-1}}, B = \frac{1}{1 - \alpha\beta^{-1}}$$

or

$$A = \frac{\alpha}{\alpha - \beta}, B = \frac{-\beta}{\alpha - \beta}$$

Now,

$$\frac{\alpha}{\alpha - \beta} \xrightarrow{z^{-1}} \alpha \cdot \alpha^n \cdot u[n] = \alpha^{n+1} \cdot u[n]$$

and

$$\frac{\beta}{\alpha - \beta} \xrightarrow{z^{-1}} \beta \cdot \beta^n \cdot u[n] = \beta^{n+1} \cdot u[n]$$

Therefore,

$$Y(z) \xrightarrow{z^{-1}} y[n]$$

and

$$y[n] = \frac{(\alpha^{n+1} - \beta^{n+1})}{(\alpha - \beta)} u[n]$$

or,

$$y[n] = \frac{(\beta^{n+1} - \alpha^{n+1})}{(\beta - \alpha)} u[n]$$

Method 2:

By convolution property of *z-transform*

$$Y(z) = \frac{1}{(1 - \alpha z^{-1})(1 - \beta z^{-1})} = \frac{1}{(1 - \alpha z^{-1})} \cdot \frac{1}{(1 - \beta z^{-1})}$$

For

$$\frac{1}{(1 - \alpha z^{-1})(1 - \beta z^{-1})}; \text{ ROC : } |z| > |\beta|$$

For

$$\frac{1}{(1 - \alpha z^{-1})}; \text{ ROC : } |z| > |\alpha|$$

For

$$\frac{1}{(1 - \beta z^{-1})}; \text{ ROC : } |z| > |\beta|$$

So,

$$\frac{1}{(1 - \alpha z^{-1})}; \text{ ROC : } |z| > |\alpha| \xrightarrow{z^{-1}} \alpha^n \cdot u[n]$$

7. chapter

Similarly,

$$\frac{1}{(1 - \beta z^{-1})}; \text{ ROC : } |z| > |\beta| \xrightarrow{z^{-1}} \beta^n \cdot u[n]$$

Therefore,

$$\frac{1}{(1 - \alpha z^{-1})(1 - \beta z^{-1})}; \text{ ROC : } |z| > |\beta| \xrightarrow{z^{-1}} \sum_{k=-\infty}^{\infty} \alpha^k \cdot u[k] \cdot \beta^{n-k} u[n-k]$$

$u[k] = 1$, only for $n \geq 0$

$u[n-k] = 1$, only for $n-k \geq 0$

$u[k] \cdot u[n-k] = 1$, only for $0 \leq k \leq n$

$$\Rightarrow y[n] = \left(\sum_{k=0}^n \alpha^k \beta^{n-k} \right)$$

Here $u[n]$ comes from the fact that only for $n \geq 0$, can $0 \leq k \leq n$ be a non-trivial interval.

$$\begin{aligned} y[n] &= \beta^n \cdot u[n] \sum_{k=0}^n \alpha^k \beta^{-k} \\ y[n] &= \beta^n \cdot u[n] \sum_{k=0}^n (\alpha \beta^{-1})^k \\ y[n] &= \beta^n \cdot u[n] \left(\frac{1 - (\alpha \beta^{-1})^{n+1}}{1 - \alpha \beta^{-1}} \right) \\ y[n] &= \beta^{n+1} \cdot u[n] \left(\frac{1 - (\alpha \beta^{-1})^{n+1}}{\beta - \alpha} \right) \\ y[n] &= \left(\frac{\beta^{n+1} - \alpha^{n+1}}{\beta - \alpha} \right) \end{aligned}$$

The results from both the methods are same.

7.9.3. ***z-transforms with repeated factors***

Let us take an example.

$$H(z) = \frac{1}{(1 - \frac{1}{2}z^{-1})^2 \cdot (1 - \frac{1}{3}z^{-1})} : \text{ ROC : } |z| > \frac{1}{2}$$

and we need to find the inverse *z-transform*.

So the first step is to decompose $H(z)$ into partial fractions.

$$\begin{aligned} H(z) &= \frac{A_{11}}{1 - \frac{1}{2}z^{-1}} + \frac{A_{21}}{(1 - \frac{1}{2}z^{-1})^2} + \frac{A_2}{(1 - \frac{1}{3}z^{-1})} \\ &= \frac{1}{(1 - \frac{1}{2}z^{-1})^2 \cdot (1 - \frac{1}{3}z^{-1})} \end{aligned}$$

Here we see that the repeated factor occur in all its powers.

Now to find A_{21} multiply both sides by $(1 - \frac{1}{2}z^{-1})^2$ and put $z = \frac{1}{2}$.

Solving this we get, $A_{21} = 3$.

A_2 can be found by multiplying both sides by $(1 - \frac{1}{2}z^{-1})$ and put $z = \frac{1}{3}$
Solving this we get $A_2 = 4$.

To find A_{11} , solve this

$$A_{11} = (1 - \frac{1}{2}z^{-1}) \cdot \left(\frac{1}{(1 - \frac{1}{2}z^{-1})^2 \cdot (1 - \frac{1}{3}z^{-1})} - \frac{3}{(1 - \frac{1}{2}z^{-1})^2} - \frac{4}{(1 - \frac{1}{3}z^{-1})} \right)$$

$$A_{11} = -6$$

There is more generalised way to find A_{11} .

Let us say $H(z)$ is of the following form

$$H(z) = \frac{A_{11}}{1 - \frac{1}{2}z^{-1}} + \frac{A_{21}}{(1 - \frac{1}{2}z^{-1})^2} + \frac{A_2}{1 - \frac{1}{3}z^{-1}}$$

Multiply both sides by $(1 - \frac{1}{2}z^{-1})^2$, giving

$$(1 - \frac{1}{2}z^{-1})^2 \cdot H(z) = (1 - \frac{1}{2}z^{-1}) \cdot A_{11} + A_{22} + A_2 \frac{(1 - \frac{1}{2}z^{-1})^2}{(1 - \frac{1}{3}z^{-1})}$$

Differentiate both sides w.r.t. z^{-1} and put $(1 - \frac{1}{2}z^{-1}) = 0$ or $z = 2$, we get

$$-\frac{1}{2}A_{11} = \left. \frac{d}{dz^{-1}} (1 - \frac{1}{2}z^{-1}) \cdot H(z) \right|_{z=2}$$

Solving this we get, $A_{11} = -6$.

This method is more general for the higher power of the factors, for example

$$H(z) = \frac{1}{(1 - \frac{1}{2}z^{-1})^3 \cdot (\dots) \cdot (\dots)}$$

$$H(z) = \frac{A_{11}}{1 - \frac{1}{2}z^{-1}} + \frac{A_{21}}{(1 - \frac{1}{2}z^{-1})^2} + \frac{A_{31}}{(1 - \frac{1}{2}z^{-1})^3} + \text{other terms}$$

Multiply both sides by $(1 - \frac{1}{2}z^{-1})^3$, we get

$$(1 - \frac{1}{2}z^{-1})^3 \cdot H(z) = (1 - \frac{1}{2}z^{-1})^2 A_{11} + (1 - \frac{1}{2}z^{-1}) A_{21} + A_{31} + \text{other terms}$$

Now successively differentiate w.r.t. z^{-1} and put $z = 2$.

So without any differentiation we get A_{31} , with the first differentiation we would get A_{21} and with the second differentiation we would get A_{11} .

7.9.4. Differentiation w.r.t. ' z' : Time Shifting in the natural domain

Let us take $(1 - \frac{1}{2}z^{-1})^{-2}$

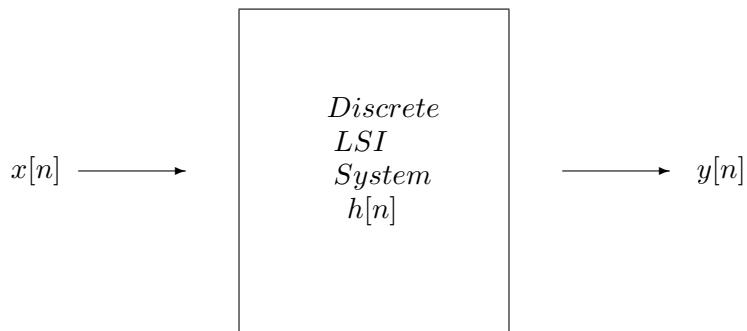
$$\frac{d}{dz} \left(\frac{1}{(1 - \frac{1}{2}z^{-1})^2} \right) = \frac{(-2)(-\frac{1}{2}z^{-1})(-1)}{(1 - \frac{1}{2}z^{-1})^3}$$

$$(-z) \frac{d}{dz} \left(\frac{1}{(1 - \frac{1}{2}z^{-1})^2} \right) = \frac{2 \cdot \frac{1}{2} \cdot z^{-1}}{(1 - \frac{1}{2}z^{-1})^3}$$

Here ' $(-z) \frac{d}{dz}$ ' factor corresponds to multiplication by ' n ' in the natural domain.

And to remove z^{-1} factor, replace ' n ' by ' $n + 1$ ' i.e. advance the sequence by 1.

7.9.5. System Function as a ratio



If the following are assumed

$$\begin{aligned} x[n] &\xrightarrow{Z} X(Z); \text{ ROC : } R_X \\ h[n] &\xrightarrow{Z} H(Z); \text{ ROC : } R_H \\ y[n] &\xrightarrow{Z} Y(Z); \text{ ROC : } R_Y \end{aligned}$$

Then,

$$y[n] = \sum_{k=-\infty}^{\infty} x[k] \cdot h[n-k]$$

And by convolution property

$$Y(z) = X(z) \cdot H(z)$$

Now the system function is

$$H(z) = \frac{Y(z)}{X(z)}$$

This ratio is independent of $X(z)$ for an LSI system.

7.9.6. Zeroes and Poles of the system

For a rational z -transform,

the factors for which denominator is equal to zero are called the 'poles' of the system/sequence. And

the factors for which the numerator is zero are called the 'zeroes' of the system/sequence.

7.9.7. General process for inversion of rational z-transform

A rational *z-transform* in z or z^{-1} are both equivalent.

For example $\frac{1}{(1-\frac{1}{2}z^{-1})}$ and $\frac{z}{(z-\frac{1}{2})}$ are *z-transforms* in z^{-1} and z respectively and are equal. So we can work with whichever expression we find more comfort with. Here we will work with z^{-1} .

Let us solve an example

$$H(z) = \frac{z^2 - z}{(1 - \frac{1}{2}z^{-1})}$$

We need to pull out the power of z in the numerator.

$$H(z) = z^2 \frac{(1 - z^{-1})}{1 - \frac{1}{2}z^{-1}}$$

z^2 implies that the sequence is shifted by two samples in the backward direction in the natural domain. This can be taken care of at the end of the inversion.

We now focus on the rational function of z^{-1} .

If degree of numerator \geq degree of denominator, do a long division of the numerator w.r.t. denominator.

The quotient will give a finite series and the remainder will be the numerator of the proper fraction.

$$\frac{(1 - z^{-1})}{1 - \frac{1}{2}z^{-1}} = 2 + \frac{(-1)}{(1 - \frac{1}{2}z^{-1})}$$

In general, a finite series can be combined with the initial power of ' z' . Suppose we have a finite series

$$a_0 + a_1 z^{-1} + a_2 z^{-2} + \dots$$

And we have an initial factor of z^D . So we can write the complete series as

$$S(z) = a_0 z^D + a_1 z^{D-1} + a_2 z^{D-2} + \dots$$

and this can be inverted easily

$$s[n] = a_0 \delta[n + D] + a_1 \delta[n + (D - 1)] + \delta[n + (D - 2)] + \dots$$

Hence, we get a train of discrete impulses and $s[n]$ is a finite series.

If we have a pole at $z = \alpha$ repeated M-fold.

$$\frac{(\dots)}{(1 - \alpha z^{-1})^M \cdot (\dots)}$$

This contributes a 'POLYEX' term in ' n ' of the following form

$$\Rightarrow (\text{Polynomial in } n \text{ of degree } M - 1) \cdot \alpha^n$$

7. chapter

This 'polyex' term is associated with either $u[n]$ or $u[-n]$ or an appropriately shifted unit-step ($u[n \pm D]$).

This choice between $u[n]$ or $u[-n]$ comes from whether the sequence is left-sided or right-sided.

Now let us solve the above example

$$H(z) = z^2 \frac{(1 - z^{-1})}{1 - \frac{1}{2}z^{-1}} = z^2 \cdot \left(2 + \frac{(-1)}{\left(1 - \frac{1}{2}z^{-1}\right)}\right) = 2z^2 - z^2 \frac{1}{\left(1 - \frac{1}{2}z^{-1}\right)}$$

So, by invoking the linearity property of z -transform, we can easily deal with the finite series and the rational proper fraction separately

$$\begin{aligned} 2z^2 &\xrightarrow{z^{-1}} 2 \cdot \delta[n+2] \\ z^2 \frac{1}{\left(1 - \frac{1}{2}z^{-1}\right)}; \text{ ROC : } |z| > \frac{1}{2} &\xrightarrow{z^{-1}} \left(\frac{1}{2}\right)^{n+2} \cdot u[n+2] \\ \text{or} \\ z^2 \frac{1}{\left(1 - \frac{1}{2}z^{-1}\right)}; \text{ ROC : } |z| < \frac{1}{2} &\xrightarrow{z^{-1}} \left(\frac{1}{2}\right)^{n+2} \cdot u[-n+1] \end{aligned}$$

Therefore,

$$h[n] = \begin{cases} 2 \cdot \delta[n+2] + \left(\frac{1}{2}\right)^{n+2} \cdot u[n+2] & : \text{For ROC : } |z| > \frac{1}{2} \\ 2 \cdot \delta[n+2] - \left(\frac{1}{2}\right)^{n+2} \cdot u[-n+1] & : \text{For ROC : } |z| < \frac{1}{2} \end{cases}$$

7.10. Module 4: Lecture 10

**Looking at rational Laplace transform and z transform together
And**

Determining Causality of a LTI system from the system function

7.10.1. Introduction

Till now, we have dealt with Laplace transform and z transform independently. Here, we take a look at them together and compare the procedures to get the inverse of either of them given the transforms and the regions of convergence.

Also, in this lecture we define the system function for continuous time and discrete time linear shift invariant systems. Further, we will learn to examine the causality of a LTI system by looking at its system function.

7.10.2. System function

The system function of a continuous independent variable LTI system is defined to be the Laplace transform of its impulse response.

Whereas, the system function of a discrete independent variable LTI system is defined to be the z transform of its impulse response.

7.10.3. Looking at rational Laplace transform and z transform together

Both the rational Laplace transform and the rational z transform are associated with poles and zeroes.

Any rational function $p(x)/q(x)$ (where both $p(x)$ and $q(x)$ are polynomials) can be written in the following decomposed form:

$$\frac{p(x)}{q(x)} = \text{Quotient} + \frac{\text{Remainder}}{\text{Denominator}}$$

Quotient part

The quotient part is a finite series in s or z depending on the transform being dealt with. For example,

$$a_0 + a_1s + a_2s^2 + \dots + a_ns^n + a_{-1}s^{-1} + \dots + a_{-n}s^{-n}$$

Laplace transform

If the system function of a continuous independent variable LTI system has a quotient part which is anything other than a constant, the system immediately becomes unstable. (Since, e.g. an s term in the transform corresponds to differentiation and a $1/s$ term corresponds to integration).

Z transform

The quotient part of the system function of a discrete independent variable system contributes to the impulse response of the system. By itself, it can never change the stability or the instability of the system.

Comparing the effect of quotient part of the system function on the stability of the system

The quotient part of the system function of a continuous independent variable and discrete independent variable LTI system present a contrasting behavior.

The presence of a non constant term in the quotient part of the system function of continuous independent variable LTI system makes the system unstable.

Whereas, the above statement is not true for the quotient part of the z-transform. This difference is to be noted.

Remainder part

Now consider the (Remainder/Denominator) part. We can express this in terms of partial fractions. Each distinct pole which appears in the partial fraction expansion gives a “polyex” term (poly-polynomial, ex-exponential).

7. chapter

Laplace transform

In Laplace transform, a factor like $1/(s - a)^M$ with other factors in the denominator, gives a "polyex" term of the form:

$$e^{at} * (\text{polynomial in } t \text{ of degree } M - 1) * (u(t) \text{ or } u(-t))$$

Note: $u(t)$ or $u(-t)$ in the above expression depends on the region of convergence (ROC) of the Laplace transform.

We identify vertical line passing through $s = a$.

If the ROC is to the left of it, then we have left-sided solution. Therefore, we multiply by $u(-t)$.

If the ROC is to the right of it, then we have right sided solution. Therefore, we multiply by $u(t)$ as is already seen.

Z transform

Suppose the z transform has a term of the form $1/(1 - az^{-1})^M$ multiplied by other factors in numerator and denominator. On inversion this gives a "polyex" term of the form:

$$a^n * (\text{a polynomial in } n \text{ of degree } M-1) * (u[n] \text{ or } u[-n] \text{ or their shifted version})$$

where the choice between $u[n]$ and $u[-n]$ is based on whether the ROC is to the exterior or the interior of the circle $|z|=|a|$.

If the the ROC lies to the exterior of this circle, we have the right sided sequence $(u[n])$.

If the ROC lies inside this circle we have the left sided sequence $(u[-n])$.

An important point to note is that the region of convergence can never pass through a pole. This is because a pole is a point of singularity and so z transform or Laplace transform can never converge at a pole.

Summary

We identify the poles for the rational Laplace transform or the rational z transform as the case may be.

For the Laplace transform, we draw vertical lines to these poles and the possible regions of convergence are interspersed in between these vertical lines.

In the z transform, we draw circles centered at the origin passing through the poles and the possible regions of convergence are discs between the circles, i.e. the intervening regions between these circles.

7.10.4. Illustration

Consider the following:

$$H(z) = \frac{(1 - \frac{1}{4}z^{-1})}{(1 - 2z^{-1})^2(1 - \frac{1}{3}z^{-1})}$$

7.10. Module 4: Lecture 10 Looking at rational Laplace transform and z transform together And Determining ROC

Thus we have a zero at $z = \frac{1}{4}$ and poles at $z = 2$ with multiplicity 2 and at $z = \frac{1}{3}$.

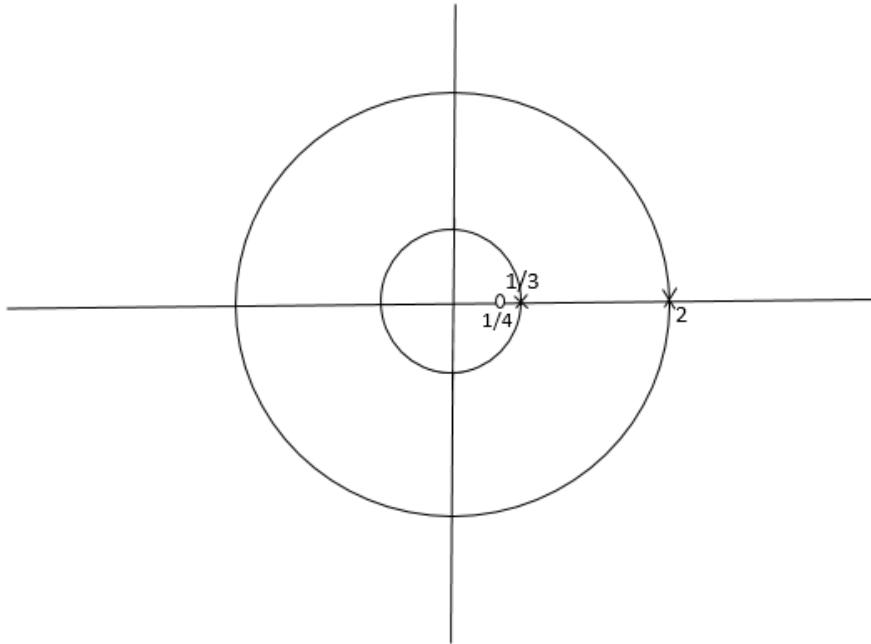
The degree of numerator is less than the degree of the denominator and we can straight away make partial fraction expansion and invert.

The inverse will be of the form:

$$[(a_{10} + a_{11}n) * 2^n * (u[n] \text{ or } u[-n])] + [b_0 * (\frac{1}{3})^n * (u[n] \text{ or } u[-n])]$$

We select $u[n]$ or $u[-n]$ depending on the ROC.

The pole zero plot is as shown above. We express poles by putting crosses and



zeroes by circles.

There are 3 possible ROCs. One is inside the circle of radius $1/3$ (ROC1). The second one is between the 2 circles (ROC2) and the third ROC is outside both the circles (ROC3).

Now if an ROC is to the exterior of the circle corresponding to a pole, we choose the right sided sequence so $u[n]$ is the choice. If ROC is in the interior of the circle corresponding to a pole, we choose the left sided sequence so $u[-n]$ is the choice.

Another point to note is that there may be some loose factors in the expression of $H(z)$ such as z^2 . To take care of that, we need to shift $u[n]$, in this case to $u[n+2]$. In case of z^{-1} , we would use $u[n-1]$.

ROC1:

ROC is to the interior of circles corresponding to both the poles so $u[-n]$ for both. So we have

$$(a_{10} + a_{11}n) * 2^n * u[-n] + b_0 * (\frac{1}{3})^n * u[-n]$$

7. chapter

ROC2:

ROC is to the exterior of the circle corresponding to the pole at $1/3$ and interior of the circle corresponding to the pole at 2 . So, we get

$$(a_{10} + a_{11}n) * 2^n * u[-n] + b_0 * \left(\frac{1}{3}\right)^n * u[n]$$

ROC3:

ROC is to the exterior of both the poles. So, in this case we get,

$$(a_{10} + a_{11}n) * 2^n * u[n] + b_0 * \left(\frac{1}{3}\right)^n * u[n]$$

Conclusion:

We saw the differences and similarities in the effect of the quotient and the remainder parts of Laplace and z transforms. We also saw how the regions of convergence affect the inverse transforms of both the Laplace and z transforms.

7.10.5. Determining the causality of a LTI system from its system function

The system function of a LTI system is as complete a representation of the system as is the impulse response. Therefore, given a system function we should be able to tell if the system is causal or not. In this section, we learn methods to comment on the causality of a system given its system function for both continuous independent variable and discrete independent variable LTI systems.

Relation between system function and causality of the LTI system for continuous independent variable system.

Causal system

A continuous independent variable system is causal if its impulse response $h(t)$ is 0 for all $t < 0$.

Thus, the system function(if it exists) of a causal continuous independent variable LTI system is given by

$$H(s) = \int_0^{\infty} h(t) e^{-st} dt$$

Since s is a complex quantity, we can substitute s by $(\Sigma + j\Omega)$ where Σ and Ω are real numbers.

Thus,

$$\begin{aligned} H(s) &= \int_0^{\infty} h(t) e^{-(\Sigma+j\Omega)t} dt \\ H(s) &= \int_0^{\infty} h(t) e^{-\Sigma t} e^{-j\Omega t} dt \end{aligned}$$

We now need to worry about the ROC of the above Laplace transform.

Let us look at the integrand in the above integral.

The integrand is

$$h(t)e^{-\Sigma t} e^{-j\Omega t}$$

The modulus of the integrand is therefore given by,

$$\text{Modulus of integrand} = |h(t)e^{-\Sigma t} e^{-j\Omega t}|$$

$$= |h(t)e^{-\Sigma t}| \quad (\text{since } |e^{-j\Omega t}| = 1)$$

When $\Sigma \rightarrow \infty$, the modulus of the integrand, i.e., $|h(t)e^{-\Sigma t}| \rightarrow 0$ for all $t > 0$.

The fact that we consider only positive t is a result of $h(t)$ being zero for all $t < 0$ (or in other words, the system being causal).

Since, the integrand $|h(t)e^{-\Sigma t}| \rightarrow 0$ for all $t > 0$ when $\Sigma \rightarrow \infty$, the Laplace transform exists (converges) at the contour $\Sigma \rightarrow \infty$.

Therefore, the contour $\Sigma \rightarrow \infty$ is included in the ROC of the Laplace transform of a causal system.

Non-causal system

If the system were non-causal, $h(t)$ would have been non-zero for some negative t . This would lead to the system function

$$H(s) = \int_{-\infty}^{\infty} h(t) e^{-\Sigma t} e^{-j\Omega t} dt$$

Now, if $\Sigma \rightarrow \infty$, $|h(t)e^{-\Sigma t}| \rightarrow \infty$, for negative t .

Therefore, the integral also diverges at the contour $\Sigma \rightarrow \infty$ and thus it is NOT included in the ROC.

Conclusion

A continuous independent variable LTI system is causal if and only if the contour $\Sigma \rightarrow \infty$ lies in the ROC of the system function.

Relation between system function and causality of the LTI system for discrete independent variable system

Causal system

We know that the system function of a discrete independent variable LTI system is the z transform (if it exists) of the impulse response.

A continuous independent variable system is causal if its impulse response $h[n]$ is 0 for all $n < 0$.

Thus, the system function of a causal discrete independent variable LTI system is given by

$$H(z) = \sum_0^{\infty} h[n]z^{-n}$$

We now need to worry about the ROC of the above z transform.

Let us look at the summand in the above summation.

The summand is

$$h[n]z^{-n}$$

The modulus of the summand is therefore given by,

$$\begin{aligned} \text{Modulus of summand} &= |h[n]z^{-n}| \\ &= |h[n]| |z|^{-n} \end{aligned}$$

7. chapter

The modulus of the summand, $|h[n]| |z|^{-n} \rightarrow 0$ for all $n > 0$ and, is equal to $|h[n]|$ for $n = 0$.

Thus, the z transform also exists (converges) at the contour $|z| \rightarrow \infty$

The fact that we consider only non-negative n is a result of $h[n]$ being zero for all $n < 0$ (or in other words, the system being causal).

Therefore, the contour $|z| \rightarrow \infty$ is included in the ROC of the z transform of a causal system. Loosely speaking this contour is the outermost circle in the z plane.

Non causal system

If the system were non-causal, $h[n]$ would have been non-zero for some negative n. This would lead to the system function

$$H(z) = \sum_{-\infty}^{\infty} h[n]z^{-n}$$

Now, if $|z| \rightarrow \infty$, the summand $|h[n]| |z|^{-n} \rightarrow \infty$, for negative n. Therefore, the summation also diverges at the contour $|z| \rightarrow \infty$ and thus it is NOT included in the ROC.

Conclusion

A discrete independent variable LTI system is causal if and only if the contour $|z| \rightarrow \infty$ lies in the ROC of the system function.

7.11. Module 4: Lecture 11

Determining system stability using Laplace transforms

7.11.1. Introduction

In the previous session, we looked at the causality of a linear shift invariant system by looking at the system function. We now realize the importance of the region of convergence, which had the information about causality. In this session, we will study the property of stability of a linear shift invariant system by just looking at the system function and the importance of region of convergence in determining stability.

Throughout the process, we will look at both continuous and discrete variable systems, that is both the Laplace transform (a function of s) and Z transform (a function of z). In this lecture, we will focus only on the continuous variable systems.

We also look at the basic disqualification about the stability of the system by analyzing the system function. We will see in this lecture that if there is differentiator or integrator in the system function, there will be immediate instability.

7.11.2. Significance of Region of Convergence in Stability

Assume Linear Shift Time Invariant System which has an Laplace Transformable impulse response $h(t)$ with the region of convergence R_X .

$$h(t) \xrightarrow{\mathcal{L}} H(s) ; R_X.$$

Assume that $H(s)$ is rational. Thus, we are not looking at all continuous variable Laplace transformable impulse responses because working with general impulse responses will be tougher and rational functions itself is a big exhaustive set. Now, the expression will not contain the information about stability so we will have to identify the region of convergence also.

For example,

$H(s) = \frac{1}{s+2}$ with pole $s = -2$. Observe that it is common to two situations. There will be two possible regions of convergence $\text{Re}(s) > -2$ and $\text{Re}(s) < -2$.

For ROC $\text{Re}(s) > -2$,

The inverse Laplace transform of the expression is $e^{-2t}u(t)$ which is a decaying exponential starting from 1 at $t = 0$ and tends to 0 at $t = \infty$.

For ROC $\text{Re}(s) < -2$,

The inverse Laplace transform of the expression is $-e^{-2t}u(-t)$ which is a growing exponential starting from -1 at $t = 0$ and tends to $-\infty$ at $t = -\infty$.

For stability we will want to have the absolute integral of the impulse response existing,

$$\int_0^\infty |h(t)|dt < \infty$$

For $\text{Re}(s) < -2$,

The absolute integral is $\int_0^\infty |e^{-2t}u(t)|dt$ which converges thus it is a stable system.

However for $\text{Re}(s) < -2$,

The absolute integral is $\int_0^\infty |-e^{-2t}u(-t)|dt$ which diverges and hence it is an unstable system.

Though we have taken the same expression for the impulse response but we had two different conclusions for stability because of the different regions of convergence which clearly indicates the importance of region of convergence.

7.11.3. Poles and the Region of Convergence

We need to analyse the reason behind the two different conclusions in regard to stability for the same impulse function. Poles play an important role in determining the stability of the system. For a rational system, to define the regions of convergence, we just need to draw vertical lines on these poles and the regions in between these vertical lines are the possible regions of convergence.

In the above example we have just one pole so we either go to the left of the pole or to the right. In general, even for multiple poles, the region of convergence lies only to the left of the pole or to the right but not on the pole. This is also true for multiple poles.

7. chapter

If we focus on a given pole, we get a left-sided "polyex" term for region of convergence to the left and similarly a corresponding right-sided "polyex" term. The term "polyex" is emphasized because we can have repeated poles.

7.11.4. Analytic Explanation

Lets consider the influence of only one pole on stability of a system. As we had seen earlier, for stability we need to have the impulse function to be absolutely integrable. For a rational system $H(s)$, $h(t)$ is sum of polyex terms one corresponding to each distinct pole and possibly "singularity function". The rational function $H(s)$ can be written in numerator and denominator form as,

$$\begin{aligned} H(s) &= \frac{N(s)}{D(s)} \\ &= Quo(s) + \frac{Rem(s)}{1 + \tilde{D}(s)} \end{aligned}$$

where $Quo(s)$ is the quotient polynomial and $Rem(s)$ is the remainder polynomial when $N(s)$ is divided by $D(s)$. Both $N(s)$ and $D(s)$ can be expressed as a finite polynomial series in s. $\tilde{D}(s)$ is the denominator when the constant term in the denominator is 1, i.e.there is a constant term.

Now the degree of the remainder polynomial will be less than the degree of denominator. If the numerator degree is greater than the denominator degree. That is if the quotient is a polynomial of degree other than 0, it means there are differentiators or integrators present. There is a differentiation or double differentiation operation present in the system. The moment we have a double differentiation operation, instability is implicit. If the degree($Quo(s)$) ≥ 1 , the system is immediately unstable.

This was just the beginning of stability check when we can conclude about stability even without evaluating the remainder. sectionEffect of Differentiator and Integrator Suppose we have a system function and then we do a long division upon it and write it in the form of quotient remainder and denominator. If the quotient part of system functions includes differentiator or integrator, there is immediate instability in the system.

Let us see why differentiator and integrator are inherently unstable. We will prove this by counterexample. Consider a differentior such that $y(t)$ is the output for input $x(t)$.

$$x(t) \xrightarrow{y(t)=\frac{dx(t)}{dt}} y(t) .$$

Assume that $x(t) = \sin(t^2)$. The output $y(t)$ of the system is equal to $2t\cos(t^2)$. We note that $|x(t)|$ is less than one for all possible t . But in contrast $|y(t)|$ grows without bound in t . Thus a bounded input produced an unbounded output and hence this system is unstable.

Let us look a similar proof for integrator why they are inherently unstable. Consider a running integrator such that $y(t)$ is the output for input $x(t)$.

$$x(t) \xrightarrow{y(t)=\int_{-\infty}^t x(\lambda)d\lambda} y(t) .$$

Assume that $x(t)$ is the unit step function $u(t)$. The output $y(t)$ turns out to be $tu(t)$. We note that $x(t)$ is a bounded input whereas output $y(t)$ is unbounded. Thus a bounded input produced an unbounded output and hence this system is unstable.

7.11.5. System function of Differentiator and Integrator

The system function of differentiator is s . The system function of running integrator is $\frac{1}{s}$. We can also have cascaded differentiator or integrator. Suppose if two differentiator are cascaded then the corresponding system function is s^2 . Similarly if two integrator are cascaded then the corresponding system function is $\frac{1}{s^2}$. These type of cascaded systems can be immediately ruled out as unstable system.

7.11.6. Further analysis of System Stability

Let us assume that system does not have differentiator or integrator and look at the non-trivial case. This assumption implies that degree of numerator is less than that of denominator. We can decompose such a system function into partial function. We can invert each partial fraction term, corresponding to each distinct pole. We do not call repeated pole as distinct pole. We can club all the terms corresponding to a given pole and create a polyex term. The exponential parameter comes from the pole whereas polynomial parameters come from the coefficients in the partial function.

Now we will look at any particular polyex term. The region of convergence is either to the left of the pole or right of the pole. From left sided term we cannot infer that it is growing or decaying. Growth and decay depends upon the real part of the exponential term.

Focus on one of the polyex term.

Lemma : The absolute integral of this polyex term diverges if the exponential is growing and converges if the exponential is decaying.

The polynomial term has no role in this. So, it does not matter if the pole is repeated. We have to only consider if the pole contributes to a growing exponential or decaying exponential.

Now suppose if we have a growing exponential coming from one of the polyex term. Is it possible that one of the other polyex term can overcome that growth such that the overall impulse response of the system is in control. But this is not possible.

Lemma : The polyex are linearly independent.

One exponentially growing polyex term makes the system impulse response non-integrable and unstable. A polyex term would be exponential if one of these holds.

The pole is of positive real part and the ROC is to the right of the pole, or

The pole is negative of real part and the ROC is to the left of the pole.

Suppose if a pole is to the right of the imaginary axis, the region of convergence should be in the left and if a pole is to the left of the imaginary axis, the region of convergence should be in the right for the system stability. In either of the case imaginary axis is included in the region of the convergence for stability. So the gist is, look at the imaginary axis and if it is included in the region of convergence the system is stable.

7.11.7. Conclusion

In this lecture, we looked at the importance of Radius of Convergence in determining the stability of a system and the various conditions associated with it. We also considered a rational impulse response with finite polynomials of numerator and denominator to chalk out some conditions for stability with regard to them. In the next session, we will look into the conditions of stability or instability if we are not successful in concluding directly.

Next, we looked how differentiators and integrator affects system stability. We considered non-trivial cases of system function and commented upon their stability. We also looked at the importance of imaginary axis while determining system stability.

7.12. Module 4: Lecture 12

Stability in Laplace and Z-Domains

7.12.1. Introduction

We developed an understanding of the stability of rational systems in the continuous independent variable. We saw that how Region of Convergence (ROC) of Laplace transform of continuous time system helps in determining whether the system is stable or not.

In this lecture we first look at case when poles are at imaginary axis, what does it contribute to the impulse response and we also discuss when is a rational LSI (Linear Shift Invariant) system both causal and stable.

Next we talk about what happens when the system, described in the discrete time domain is linear, shift-invariant and the system impulse response has a Z-transform.

7.12.2. Poles on the imaginary axis

If we have poles on imaginary axis then obviously imaginary axis cannot be in ROC and hence the system will be unstable. As we saw in previous lecture that integrator is unstable. We can relate it to the fact that it has one pole on zero (which is of course on imaginary axis) and hence the system is unstable. Now, if instead of zero we have poles at 2 complex conjugate points on imaginary axis $j\Omega_c$ and $-j\Omega_c$. If the poles are simple, then $H(s)$ will be of the form $-\frac{A}{s-j\Omega_c} + \frac{\bar{A}}{s+j\Omega_c}$ in the partial fraction expansion assuming all the coefficients in Laplace transform are real. Now, the above terms will contribute $Ae^{j\Omega_ct} + \bar{A}e^{-j\Omega_ct}$ multiplied with $u(t)$ or $u(-t)$ to the impulse response. This term can then be written as $2\operatorname{Re}(Ae^{j\Omega_ct}) = 2\operatorname{Re}(Ae^{j\Omega_ct}) + 2A_0\cos(\Omega_ct + \phi + 0)$ where A_0 is real. SO, we get a sinusoidal component due to complex conjugate poles on imaginary axis.

One can show that for an LSI system of the form we described above, if we give a bounded input $x(t) = \cos(\Omega_ct).u(t)$ as input to the system then we get a polyex term at the output of the system- $y(t) = (Bt+c)\cos(\Omega_ct + \phi_1)$ for some constants B and C . The polynomial term in the above polyex term, i.e., $Bt+C$ is what makes the output unbounded and hence the system unstable for nonzero B . Note that input is of same frequency as of the filter impulse response. This phenomenon where same frequency input is giving unbounded output is "Resonance" effect. For an integrator

resonance will be at zero frequency input multiplied by unit step function, i.e., a DC signal multiplied by unit step function.

We can similarly show that even if we take repeated poles on imaginary axis instead of simple poles we will still get unbounded output for some bounded input and the system will be unstable. So, it essentially completes the argument that poles on imaginary axis makes the system unstable.

7.12.3. Causality and Stability of Rational LSI system

Previously, we saw that condition for causality of rational LSI system is that its ROC must include $\text{Re}(s) \rightarrow \infty$, and for stability ROC must include imaginary axis $\text{Re}(s) = 0$. The above conditions simply imply that for rational LSI system to be both stable and causal, no pole of the system should be in right half plane. So, a rational LSI system is causal and stable if and only if all of its poles lie strictly in left half plane. Next we would look at conditions for causality and stability of rational systems in case of discrete independent variable.

7.12.4. Stability analysis in Z plane

What is in the Z transform that determines the stability of the system? For causality we have already analysed that the extreme contour in the Z plane i.e. $|z| \rightarrow \infty$ should be included in the region of convergence in the case of causality. Now what is the condition for stability? Can we look at the ROC and come to a conclusion about the stability or otherwise?

Now let's go to our point of discussion. We are going to deal with the discrete independent variable system. We have an input $x[n]$ and corresponding output $y[n]$; assume the system is linear, shift-invariant and also rational i.e. the system impulse response $h[n]$ has a Z-transform $H(z)$ which is a rational function of z .

$$h[n] \xrightarrow{Z} H(z) \quad \text{ROC : } R_H$$

Now what do we need to convince ourselves that this R_H is central to the stability of the system? Let's take an example to do that.

Let $H(z) = \frac{1}{1 - \frac{1}{2}z^{-1}}$. Here, the pole is at $Z = 1/2$; with that we have two possible regions of convergence. One is inside the circle $|z| = 1/2$ and the other is outside the circle in the z-plane. Let's invert the Z transform in each ROC.

1. inside the circle : $h[n] = -\left(\frac{1}{2}\right)^n u[-n - 1]$
2. outside the circle : $h[n] = \left(\frac{1}{2}\right)^n u[n]$

In the first case for the ROC inner part of circle $|z| = 1/2$ the inverse Z transform is not absolutely summable so the system is unstable whereas in the second case for the ROC outer portion of the circle $|z| = 1/2$ the inverse Z transform is absolutely summable and so the system is stable.

So essentially we have the same expression for $H(z)$ but for two different regions of convergence we get two different systems; one is stable and the other one is unstable. This substantiates our argument that it is region of convergence that

7. chapter

determines the stability or instability of the system. Of course, the expression has an indirect role. So let's see that indirect role by another example.

Let $H(z) = \frac{1}{1-2z^{-1}}$. Here, the pole is at $|z| = 2$. With that we have two possible regions of convergence. One is inside the circle $|z| = 2$ and the other is outside that circle. Let's invert the Z transform in each ROC.

1. inside the circle : $h[n] = -(2)^n u[-n - 1]$
2. outside the circle : $h[n] = (2)^n u[n]$

Here, for the first case $h[n]$ is absolutely summable hence the system is stable whereas in the second case $h[n]$ diverges for higher values of n so it is not absolutely summable and so the system is not stable.

Here, again we have the same expression for $H(z)$ but for two different regions of convergence we get two different systems : one is stable and the other one is unstable. But now the roles of the regions of convergence reversed. The difference was in the pole of $H(z)$. In the first expression we had the pole at $z = 1/2 (< 1)$ and in the second expression at $z = 2 (> 1)$. In the case where the ROC was the interior part of the circle with centre at $z = 0$ and radius equal to the magnitude of corresponding pole we got a left sided signal which is absolutely summable in the second case but not in the first case. Similarly when the ROC was exterior part of the same circle in both cases we got a right sided signal which was absolutely summable in the first case and is not absolutely summable in the second case.

In either case we have a polyex term like before. Here the poles are simple, so the poly part is trivial i.e. it is just a constant. The exponential part is nontrivial, the left sided exponential is decaying if the pole is exterior with a magnitude greater than one, so remember it's the magnitude of the pole which is central. That's not surprising because, the regions of convergence are essentially between poles, so you see what kinds of regions of convergence we have in the z- plane, we are familiar with that.

How do you determine all the possible regions of convergence? You draw circles if you're talking about a rational system. We begin from z equal to zero and move outwards, identify all clean disks between these circles. A clean disk is one with no poles inside. Each such disk is a region of convergence. Now, you also know that any particular region of convergence is such that a given pole is either to the exterior of that region of convergence or to the interior. For a given region of convergence with the same expression, take a given pole. It is either to the exterior or to the interior of the ROC. Now write a partial fraction expansion, identify terms for that pole, invert terms for that pole. So you know if you focus on one pole, you can clearly invert the terms, knowing whether the pole is exterior or interior. If it is exterior to the region of convergence, it contributes left sided terms and in general if the pole is repeated you get a polyex term. The polynomial would be nontrivial, a polynomial in n . So you'd get a polyex term with degree more than just a constant, if the pole is repeated. But it would be a left sided exponential, in contrast if the pole is to the interior of the region of convergence you get a right

sided exponential with a polynomial in n multiplying it.

7.12.5. Conclusion

Magnitude of the pole is central to the analysis of stability. Any particular region of convergence is such that a given pole is either to the exterior or to the interior to the circle. If you focus on one pole, you can clearly invert the terms, knowing whether the pole is exterior or interior to the circle.

7.13. Module 4: Lecture 13

7.13.1. Introduction

In the previous lectures we have been talking about the stability of a Discrete Independent Variable System. We will be reviewing those ideas and analyze the concept of stability and causality of Discrete Independent Variable System using z-transforms. We will then try to find a correlation between z-transform and Laplace transform.

7.13.2. Stability and Causality of Discrete System

We will first discuss a general procedure to start analyzing the stability and causality of a rational system of discrete independent variables.

1. Look at one pole at a time
2. Identify its contribution to the impulse response
3. The contribution will be a *polyex* term. Where *poly* refers to the polynomial part and *ex* the exponential part.

We will now proceed by stating some lemmas and theorems in this domain.

Lemma : If any one of these *polyex* terms has a growing exponential the impulse response is not *absolutely summable*. This can be expressed as follows, if

$$\begin{aligned} h_1[n] &= P_1[n]\alpha^n \\ h_2[n] &= P_2[n]\beta^n \end{aligned}$$

where, $\alpha \neq \beta$, then there do not exist any non-zero constant pairs c_1, c_2 which satisfy $c_1h_1[n] + c_2h_2[n] = 0 \forall n$

This analysis implies that a discrete variable rational system is stable if and only if all the *polyex* terms involved have a exponential decay in the exponential part.

Now we know, for a pole with $|z_p| > 1$ to contribute a decaying exponential, it must give rise to a *left sided sequences*, i.e., the ROC must be interior of the pole. Similarly, for a pole with $|z_p| < 1$ to contribute a decaying exponential, it must give rise to a *right sided sequences*, i.e., the ROC must be exterior of the pole.

We hence notice that the unit circle is always included. This determines the general idea, that if unit circle is in the ROC then the system is stable.

7. chapter

Theorem : A rational discrete system is stable if and only if the region of convergence of its system function includes the unit circle $|z| = 1$.

For a system to be causal, we are aware that the only terms that you allow are the right-sided ones, as you wish the contour $|z| \rightarrow \infty$, to lie in the ROC. Hence, the poles must lie within the unit circle for the unit circle being include in the ROC.

Theorem : A rational discrete system is causal and stable if and only if all its pole lie within the unit circle $|z| = 1$.

7.13.3. Relationship between z-Transform and Laplace Transform

We will now review some similarities in the Laplace and z-transforms. For the Laplace transform, the critical axes are $\text{Re}(s) = 0$ and $\text{Re}(s) \rightarrow \infty$, for determining stability and causality respectively. Similarly, for the z-transform, the critical contours are $|z| = 1$ and $|z| \rightarrow \infty$, for determining stability and causality respectively.

To relate the Laplace and z-transform, we will sample a continuous time signal $x(t)$ and compare the z-transform of $x[n]$ with the Laplace transform of $x(t)$.

$$x(t) \xrightarrow{\mathcal{L}} X(s), R_X$$

Let us now sample ideally $x(t)$ with a sampling time T_S .

$$\begin{aligned} x(t) & \sum_{n=-\infty}^{\infty} \delta(t - nT_S) \\ & \sum_{n=-\infty}^{\infty} x(nT_S) \delta(t - nT_S) \end{aligned}$$

Laplace transform of the sampled signal is

$$\begin{aligned} & \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} x(nT_S) \delta(t - nT_S) e^{-st} dt \\ &= \sum_{n=-\infty}^{\infty} x(nT_S) \int_{-\infty}^{\infty} \delta(t - nT_S) e^{-st} dt \\ &= \sum_{n=-\infty}^{\infty} x(nT_S) e^{-snT_S} \end{aligned}$$

Now for the discrete sequence $x[n] = x(nT_S)$ and the z transform with $z = e^{sT_S}$, we can see that this term is exactly the same as $\sum_{n=-\infty}^{\infty} x[n]z^{-n}$.

Hence the mapping, $x[n] = x(nT_S)$ and $z = e^{sT_S}$, shows the relation between

Laplace and z-transforms.

$$\begin{aligned}|z| &= |e^{sT_s}| \\&= |e^{(\sigma+j\omega)T_s}| \\&= |e^{\sigma T_s}| \\&= |e^{Re(s)T_s}|\end{aligned}$$

This shows that the $Re(s)$ relates to the $|z|$ in the z-plane. Hence, vertical axes in s-plane corresponds to circle in z-plane.

7.13.4. Formal Inverse of z-Transform

We wish to formalize our method for finding the inverse z-transform, so that we can deal with inverting non-rational functions.

$$\begin{aligned}x[n] &\xrightarrow{z} X(z), R_X \\X(z) &= \sum_{n=-\infty}^{\infty} x[n]z^{-n}\end{aligned}$$

where z in polar form is $z = re^{j\omega}$, hence,

$$X(z) = \sum_{n=-\infty}^{\infty} x[n]r^{-n}e^{-j\omega n}$$

which is the discrete time fourier transform of $x[n]r^{-n}$. Now we can invert a DTFT.

$$\begin{aligned}y[n] &\xrightarrow{DTFT} Y(e^{j\omega}) \\Y(e^{j\omega}) &= \sum_{n=-\infty}^{\infty} y[n]e^{-j\omega n}\end{aligned}$$

where we know $e^{j\omega n}$ are orthogonal vectors. Hence,

$$y[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} Y(e^{j\omega})e^{j\omega n} d\omega$$

Now, we will apply this understanding for inverting the z-transform.

$$\begin{aligned}X(z) &= \sum_{n=-\infty}^{\infty} x[n]r^{-n}e^{-j\omega n} \\x[n]r^{-n} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} X(z)e^{j\omega n} d\omega\end{aligned}$$

we will now replace variable of integration by substituting $z = re^{j\omega}$.

$$x[n] = \frac{1}{2\pi j} \oint X(z)z^{n-1} dz$$

where the integral is over a circle of radius r and from $-\pi$ to π , where circle of r lies in the region of convergence

7.13.5. Conclusion

In this session we looked at some properties of causality and stability of discrete signals using the z-transforms. We then tried to relate the Laplace and z-transforms using sampling. We further continued our analogy by coming up with a formal expression for the inverse of the z-transform.

7.14. Module 4: Lecture 14

Signal Flow Graphs

7.14.1. Introduction

We've been doing a lot of formal constructions and formal derivations in the previous few sessions. To relate these formalisms to systems now, and getting a feel of how we can actually understand system construction, at least, for rational systems if not irrational, we will introduce the method of drawing a block diagram of a system and a way to connect it to the system description. These are called *signal flow graphs*. In a signal flow graph, we would like to realize the system using basic building components. This will be best understood if we study it using an example.

7.14.2. Continuous Independent Variable Case

Let us assume that we have a rational system with system function given by

$$H(s) = \frac{s+1}{s+2} \text{ with ROC given by } \operatorname{Re}(s) > -2$$

We can write this as

$$\begin{aligned} \frac{Y(s)}{X(s)} &= \frac{s+1}{s+2} \\ (s+2)Y(s) &= (s+1)X(s) \\ Y(s) &= -\frac{1}{2}sY(s) + \frac{1}{2}sX(s) + \frac{1}{2}X(s) \end{aligned} \tag{7.4}$$

Now, we can invert this Laplace transforms to get,

$$y(t) = -\frac{1}{2}\frac{\mathrm{d}y(t)}{\mathrm{d}t} + \frac{1}{2}\frac{\mathrm{d}x(t)}{\mathrm{d}t} + \frac{1}{2}x(t)$$

In this form, we can see what basic components are needed to generate the output.

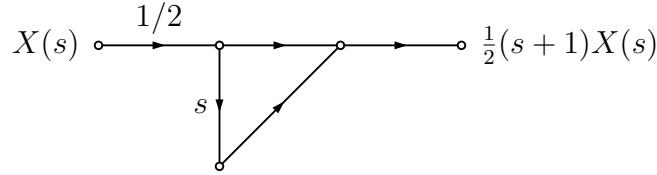
- Adder: For adding two signals (indicated by the + sign in the equation).
- Differentiator: For differentiating with respect to the independent variable (indicated by the $\frac{\mathrm{d}}{\mathrm{d}t}$ operator in the equation).
- Constant Multipliers: (Half, in this case).

Now, having identified the basic components needed, we can draw the signal flow graph. The signal flow graph consists of two elements:

1. Nodes: These contain the Laplace transformed (or Z-transformed, in case of discrete independent variable case) signals and their combinations.

2. Directed Edges: These are the transportation systems, with the multipliers embedded in them.

The meaning of these will be clear with this example. Consider the right hand side of the equation (1). The $X(s)$ dependent terms are $\frac{1}{2}sX(s)$ and $\frac{1}{2}X(s)$. Now, these will be generated by the following flow graph.

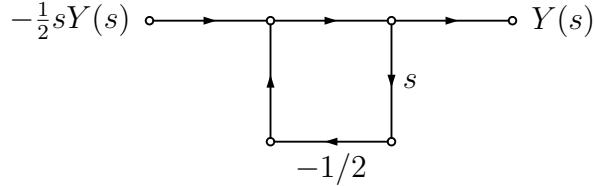


The circles are the nodes and the rays are the directed edges. The directed edges which don't have a multiplier associated are the unity multipliers, which essentially just carry the signal from one point to another. Whereas those who have a multiplier multiply the signal by the corresponding value and then direct it. This multiplier can either be constant or s dependent, as can be seen in the figure.

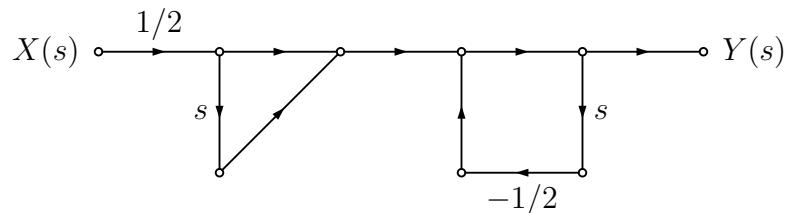
Now, suppose that $Y(s)$ is generated somewhere. We will denote this by

$$\text{---} \rightarrow Y(s)$$

Note that both its ends are $Y(s)$. Now we need to add $-\frac{1}{2}sY(s)$ to $-\frac{1}{2}(s + 1)X(s)$ to get $Y(s)$. To do that, we will have generate $\frac{1}{2}sY(s)$. It is shown as



Note that this graph has to be read from right to left. $Y(s)$ is first multiplied by half, then by s to generate $\frac{1}{2}sY(s)$ at the left. Now, to add this to the previous graph to complete the signal flow graph, we just join the two:



We now have the complete signal flow graph for the system.
As we can see in the graph, there are three types of nodes.

1. Nodes from which lines go outward, but no line comes inward. Such nodes are called *source nodes*. Here, the node containing $X(s)$ is the source node.
2. Nodes in which lines come inward, but no lines go outward. Such nodes are called *sink nodes*. Here, the node containing $Y(s)$ is the sink node.

7. chapter

3. Nodes in which some lines come inward and some go outward. Such nodes are called *intermediate nodes*. Here, all the nodes apart from the source and the sink nodes are intermediate nodes.

Notice that we have a loop here. That means you have a set of directed edges which start at the given node and reach back the same node. This happens because the output $Y(s)$ is itself dependent on $Y(s)$, which indicates that there is a recursion in the system. Loops are bound to be present in recursive systems.

Note that here we have used differentiators (multiplication by s) to realize the system. In general, it is not a good idea to use differentiation in continuous independent variable case. This is because the system function of a differentiator is $H(s) = s$. Now, the frequency response of the differentiator is given by $H(\Omega) = j\Omega$ (putting $\text{Re}(s) = 0$). We can see that the modulus of this increases linearly with Ω , that is, linearly with the frequency. Hence, a differentiator over-emphasizes higher frequencies, and the noise in the system generally exists in the higher frequencies. Hence, if a differentiator is used, the noise will get enhanced, which is not desired. Instead, we can use an integrator to realize the same system. The system function for the integrator is $H(s) = \frac{1}{s}$ and the frequency response is $H(\Omega) = \frac{1}{j\Omega}$. Now this system attenuates the higher frequencies. Hence, an integrator is preferred over a differentiator. Although the integrator over-emphasizes constant inputs, any other kind of inputs don't have a problem.

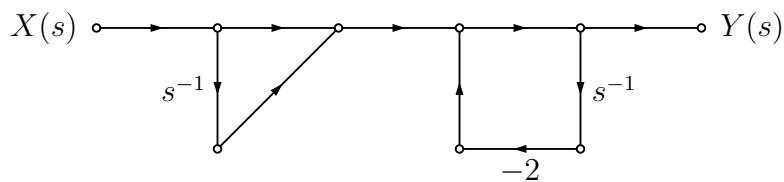
To realize this system using integrators, we write

$$H(s) = \frac{Y(s)}{X(s)} = \frac{s+1}{s+2} = \frac{s^{-1}s+1}{s^{-1}s+2} = \frac{1+s^{-1}}{1+2s^{-1}}$$

Hence, we can write

$$Y(s) = -2s^{-1}Y(s) + (1+s^{-1})X(s)$$

and we can correspondingly draw the signal flow graph as follows:



As you can see, there are more than one ways of drawing the signal flow graph for a given system. The defining equation is the same, the system function is the same, but we have different structures. Now which structure is better, which structure is worse, is a deep subject by itself. Very often we would prefer integrators instead of differentiators in continuous variable systems. However that is also not universally true. There could be some situations where you want a differentiator, for one reason or the other. Or you may want a combination of differentiators and integrators, and many structures are possible, as it is seen in this case.

Let us see how this construction translates in the discrete independent variable case.

7.14.3. Discrete Independent Variable Case

We will again understand this using an example. Let the system function of a discrete system be given by

$$H(z) = \frac{1 + z^{-1}}{1 - \frac{1}{2}z^{-1}} \text{ with ROC as } |z| > \frac{1}{2}$$

Hence, we have

$$(1 - \frac{1}{2}z^{-1})Y(z) = (1 + z^{-1})X(z)$$

We can invert the Z-transforms, giving

$$y[n] - \frac{1}{2}y[n-1] = x[n] + x[n-1]$$

We can rearrange this to get

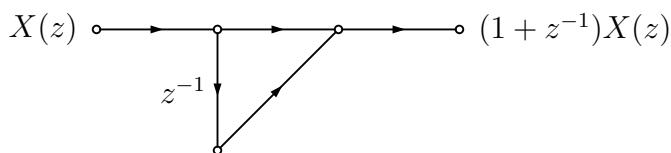
$$y[n] = \frac{1}{2}y[n-1] + x[n] + x[n-1]$$

This is an example of a Linear Constant Coefficient *Difference* Equation. Difference essentially means shifted input and output. They can be once or more delayed functions, of the form $x[n-M]$ or $y[n-M]$ for some natural number M . Higher delays can arise due to terms like z^{-2} and so on. We are considering only delayed versions and not advanced versions, to ensure causality. Now, to realize a delay of the form $x[n-M]$ for some $M > 1$, we can use unit delays M times. Hence, a unit delay is a basic building component.

Now, coming back to our example, we can write $Y(z)$ as

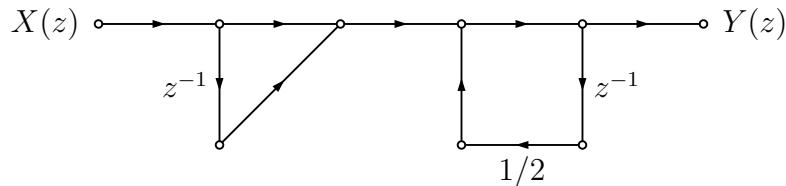
$$Y(z) = \frac{1}{2}z^{-1}Y(z) + (1 + z^{-1})X(z)$$

Analogous to the continuous independent variable case, we can draw a signal flow graph for this system. In this case, the nodes are the Z-transforms of the sequences. To generate the $X(z)$ dependent term, we can draw the following graph.



Now, as the nodes are the Z-transforms, multiplying by z^{-1} amounts to delaying the sequence by one unit.

We can similarly draw the $Y(z)$ dependent term, and finally get the signal flow graph as follows:



Now, we will look at what is the general form of output a rational system gives to an input with a rational Laplace transform or Z-transform.

7.14.4. General Form of Rational System Outputs

If the input to a rational system has a rational Laplace transform or Z-transform, then the output will also have a rational Laplace transform or Z-transform. According to the general procedure, we will segregate the poles of the output by decomposing it in to partial fractions, have a look at the region of convergence and then identify the term corresponding to each poles viz. the ROC.

Now, the output poles will be of three types, in general:

1. Input poles which are *not* system poles.
2. System poles which are *not* input poles.
3. Coincident system and input poles.

The third case can arise if the input and the system have some poles in common. The terms arising due to each of these poles are given a specific name. The terms due to the input poles are called the *forced response* terms, the ones due to the system poles are called the *natural response* terms and the ones due to the coincident poles are called the *resonant response* terms. Thus, there are three kinds of terms in general, although one of them may be absent in specific cases. So, the forced response will have terms drawn from the input, whereas the natural response will have terms drawn from the impulse response, and of course the coefficients will vary according to the partial fraction expansion.

A physical example for the resonant response is the well known fact that bridges can collapse if many people walk on it in rhythm. The bridge can be approximated to have a rational system function with simple poles. For simple poles, the corresponding terms are just sinusoidal, and won't affect the bridge. But if many people walk on it in rhythm, that may generate a strong enough input, possibly having poles at the exact same place as that of the system function of the bridge. Due to this, the system pole can become double or triple, in which case the corresponding term will be a sinusoid multiplied by time, or the square of time, etc. Hence, we will have ever increasing oscillations, which can damage the bridge, even causing it to collapse completely.

7.14.5. Conclusion

We have laid the foundation for being able to deal with at least a reasonably good class of systems and inputs. We've been able to capture the essence of the transform method, generalised also to unstable systems as you can say, in the Laplace domain and in the Z domain respectively for continuous independent variable and discrete independent variable systems.

Looking at the whole module, we have generalised the transformed domain of module two, and also to carried forth certain ideas that we had learnt in module one and module three.

This transformed domain is more general. By replacing s by $j\Omega$, we can go back to the Fourier Transform from the Laplace domain in continuous independent variable case, and similarly by replacing z by $e^{j\Omega}$ in the discrete independent variable case, to get the Discrete Time Fourier Transform. These techniques that you've learnt in all these four modules and particularly in this fourth module are used in many

branches of engineering, not just electrical engineering or mechanical engineering or systems engineering or control.