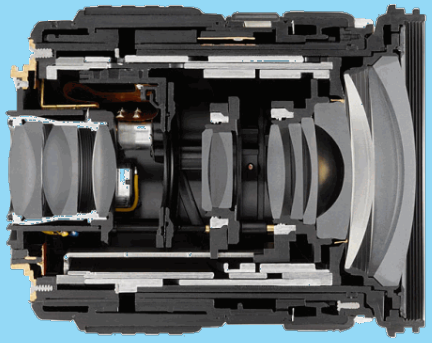
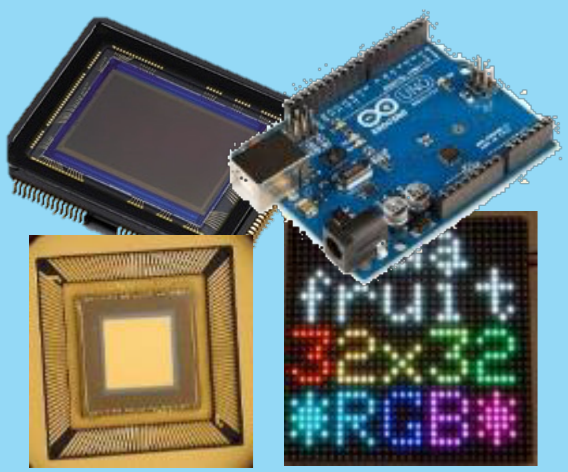


Optics

+



Sensors
&
devices

+

Signal
processing
&
algorithms

COMPUTATIONAL METHODS FOR IMAGING (AND VISION)

LECTURE 22: STATISTICAL METHODS

PROF. JOHN MURRAY-BRUCE

WHERE ARE WE

WE ARE HERE!

10	15-Mar-21	Forward Models and Inverse Problems	Linear Inversion - Inverse problems - Deconvolution and Denoising	IIP 4, Appendix E	**HW 4**	
	17-Mar-21		Intro to Regularized Inversion I - Tikhonov	IIP 5, Appendix E		
11	22-Mar-21	Regularization	Intro to Regularized Inversion II - Iterative methods - Steepest descent	IIP 6		
	24-Mar-21		Statistical methods I - ML estimation - Bayesian estimation	IIP 7.1 - 7.5		
12	29-Mar-21	Forward models and Inverse Problems II	LSV imaging systems: Forward problem - SVD - Inversion	IIP 8.1, 9, 10		
	31-Mar-21	Non-linear Regularization	Beyond L_2-regularization - Sparsity (l_0- and l_1-priors) - TV prior	SMIV 1.1 - 1.5 Papers & Handout		**HW 4**
13	5-Apr-21		Algorithms overview - ISTA/FISTA - ADMM	Papers & Handout		
	7-Apr-21	Introductory Optics	Geometrical/Ray Optics - Rays & pinhole cameras - Lenless imaging and Coded apertures	IIP 8.2, 8.3, 9.5 Papers & Handout		
14	12-Apr-21 14-Apr-21	**Spring Break (no classes)**				
15	19-Apr-21	Applications of Comp. Imaging	Looking around corners (NLOS imaging)	Papers & Handout		
	21-Apr-21		Compressive Imaging and Imaging from few photons	Papers & Handout		
16	26-Apr-21 28-Apr-21	**Group Presentations (Teams)**				
17	3-May-21	***no class**				
	5-May-21	**Final Exam: 12:30 PM - 2:30 PM**				

OUTLINE

- Maximum likelihood estimation (MLE)
- Bayesian methods (MAP)

LEARNING GOALS

- Understand MLE
- Be able to derive ML estimate under Gaussian noise
- Be able to define a prior
- Be able to interpret the role of prior as a regularizer

READING

- IIP Chapter 7

STATISTICAL METHODS

- **Statistical methods** can account for the random nature of the noise introduced due to detector/sensor/measurement imperfections.
  - Such a description means that the recorded measurement/image is the realization of a random process.
- There are two main methods:
  - **Maximum likelihood estimation (MLE) method**
  - **Bayesian Method**

UNKNOWN OBJECT IS DETERMINISTIC

MAXIMUM LIKELIHOOD (ML) METHODS

MAXIMUM LIKELIHOOD ESTIMATION

- The key distinguishing property of maximum likelihood method is that the unknown image is deterministic
- Thus can be treated as a set of deterministic parameters to be estimated
- Noise is still assumed to be a realization of a random process
  - This is useful when certain statistical properties of the noise is known, such as: the expectation value, the variance, the probability distribution
  - These can be leveraged for solving the inverse problem

MAXIMUM LIKELIHOOD ESTIMATION
HOW DOES IT WORK?

- Start with our usual discrete model of image formation:

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n},$$

- where:
  - $\boldsymbol{n} \in \mathbb{R}^M$ is unknown random noise process,
  - $\boldsymbol{y} \in \mathbb{R}^M$ is the measurement (measured image/signal)
  - $\boldsymbol{x} \in \mathbb{R}^N$ is the discretization of the unknown object
  - $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ is the discrete forward model (for the imaging system)
- The noise $\boldsymbol{n} \in \mathbb{R}^M$ is random, as such $\boldsymbol{y} \in \mathbb{R}^M$ is also a random vector.
- A **random vector** simply means that each entry of the 1D array is random:
  - We will assume that they are **independent** but have the same distribution – **independent and identically distributed (or i.i.d)**

MAXIMUM LIKELIHOOD ESTIMATION

HOW DOES IT WORK?

- Discrete model of image formation:

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n},$$

  - $\boldsymbol{n} \in \mathbb{R}^M$ is unknown random noise process,
  - $\boldsymbol{y} \in \mathbb{R}^M$ is the measurement (measured image/signal)
  - $\boldsymbol{x} \in \mathbb{R}^N$ is the discretization of the unknown object
  - $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ is the discrete forward model (for the imaging system)
- We assume that the expectation of the measurement $\boldsymbol{y} \in \mathbb{R}^M$ is:

$$E\{\boldsymbol{y}\} = \boldsymbol{A}\boldsymbol{x}$$

  - Equivalent to saying the noise process $\boldsymbol{n} \in \mathbb{R}^M$ has zero mean
- The distribution of $\boldsymbol{n} \in \mathbb{R}^M$ is assumed to be known, which means the conditional distribution of $\boldsymbol{y}$, denoted as $p_{\boldsymbol{y}|\boldsymbol{x}}(\boldsymbol{y} \,|\, \boldsymbol{x}) = p_{\boldsymbol{n}}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})$
- Remember: $\boldsymbol{x}$ is unknown and we know measurement $\boldsymbol{y}$:

$$\mathscr{L}(\boldsymbol{x}) = p_{\boldsymbol{y}|\boldsymbol{x}}(\boldsymbol{y} \,|\, \boldsymbol{x}) = p_{\boldsymbol{n}}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})$$

  - The function $\mathscr{L}(\boldsymbol{x})$ is called the **likelihood function.**

MAXIMUM LIKELIHOOD ESTIMATION

HOW DOES IT WORK?

- **Discrete model of image formation:**

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n},$$

- **Likelihood function:**

$$\mathscr{L}(\boldsymbol{x}) = p_{\boldsymbol{y}|\boldsymbol{x}}\left(\boldsymbol{y} \,|\, \boldsymbol{x}\right) = p_{\boldsymbol{n}}\left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\right)$$

- Remember: $\boldsymbol{x}$ is known, thus one can pose the question:
  - Which $\boldsymbol{x}$ from the family of candidate/possible objects is most likely to have lead to the observed noisy measurement $\boldsymbol{y}$?
  - This is the maximum likelihood (ML) estimate.
- Precisely, the **maximum likelihood estimate** of $\boldsymbol{x}$ is the object $\boldsymbol{x}_{\mathrm{ML}}$ which maximizes the likelihood $\mathscr{L}(\boldsymbol{x})$ of obtaining the observed measurement/image.

MAXIMUM LIKELIHOOD ESTIMATION
HOW DOES IT WORK?

- **Discrete model of image formation:**

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n},$$

- **Likelihood function:**

$$\mathscr{L}(\boldsymbol{x}) = p_{\boldsymbol{y}|\boldsymbol{x}}\left(\boldsymbol{y}\,|\,\boldsymbol{x}\right) = p_{\boldsymbol{n}}\left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\right)$$

- **Maximum likelihood estimate:**

$$\boldsymbol{x}_{\mathrm{ML}} = \arg\max_{\boldsymbol{x}}\ \mathscr{L}(\boldsymbol{x})$$

$$\boxed{\boldsymbol{x}_{\mathrm{ML}} = \arg\min_{\boldsymbol{x}}\ -\log(\mathscr{L}(\boldsymbol{x}))}$$

- The second optimization (i.e., the minimization problem)
  - Obtained by taking the natural logarithm and then multiplying by a minus sign (the multiplication by minus sign turns maximization into minimization).
  - It is often easier to solve (usually by using calculus) and so is the common approach.
    - Differentiating the negative log-likelihood w.r.t $\boldsymbol{x}$ and setting that to zero.

MAXIMUM LIKELIHOOD ESTIMATION

EXAMPLE 1: TOY ONE MEASUREMENT CASE

- Consider the measurement model: $y = ax + n$, where the noise n is zero mean Gaussian random variable, i.e. $n \sim \mathcal{N}(0, \sigma^2)$, y is the single measurement, the unknown is x, and a (our forward model) is a known scalar.

  Solution steps:

  1. Write the noise density function: $p_n(n) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{n^2}{2\sigma^2}}$
  2. Derive likelihood function: $\mathcal{L}(x) = p_n(y - ax) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-ax)^2}{2\sigma^2}}$
  3. Maximize $\mathcal{L}(x)$, or minimize its negative log-likelihood, i.e. minimize $-\log\left(\mathcal{L}(x)\right)$

     i) What is $\log(\mathcal{L}(x))$?

     $$\log(\mathcal{L}(x)) = \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-ax)^2}{2\sigma^2}}\right) = -\log\left(\sqrt{2\pi\sigma^2}\right) + \log\left(e^{-\frac{(y-ax)^2}{2\sigma^2}}\right) = -1/2\log(2\pi\sigma^2) - 1/(2\sigma^2)(y-ax)^2$$

     ii) Differentiate and set the negative log-likelihood to zero:

     $$-\frac{\partial}{\partial x}\log(\mathcal{L}(x)) = -(y - ax)/\sigma^2 = 0$$

     iii) Solve the resulting equation: $-(y - ax_{\mathrm{ML}})/\sigma^2 = 0 \;\Rightarrow\; y - ax_{\mathrm{ML}} = 0$, and thus $x_{\mathrm{ML}} = y/a$.

MAXIMUM LIKELIHOOD ESTIMATION

EXAMPLE 1: TOY ONE MEASUREMENT CASE

- Consider the measurement model: $y = ax + n$, where the noise n is zero mean Gaussian random variable, i.e. $n \sim \mathcal{N}(0, \sigma^2)$, y is the single measurement, the unknown is x, and a (our forward model) is a known scalar.

  Solution steps:

  1. Write the noise density function: $p_n(n) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{n^2}{2\sigma^2}}$
  2. Derive likelihood function: $\mathcal{L}(x) = p_n(y - ax) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-ax)^2}{2\sigma^2}}$
  3. Maximize $\mathcal{L}(x)$, or minimize its negative log-likelihood, i.e. minimize $-\log\left(\mathcal{L}(x)\right)$

     i)

     1

     ii)

     iii) y/a.

In a handful of simple scenarios, one can simply read off the maximizer of the likelihood function. This is one of them!

$\mathcal{L}(x)$ attains its maximum when $(y - ax)^2 = 0$, which is

MAXIMUM LIKELIHOOD ESTIMATION
EXAMPLE 1: TOY ONE MEASUREMENT CASE IN GAUSSIAN NOISE

- Consider the measurement model: $y = ax + n$, where the noise n is zero mean Gaussian random variable, i.e. $n \sim \mathcal{N}(0, \sigma^2)$, y is the single measurement, the unknown is x, and a (our forward model) is a known scalar.

  Solution steps:

  1. Write the noise density function: $p_n(n) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{n^2}{2\sigma^2}}$
  2. Derive likelihood function: $\mathscr{L}(x) = p_n(y - ax) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-ax)^2}{2\sigma^2}}$
  3. Maximize $\mathscr{L}(x)$, or minimize its negative log-likelihood, i.e. minimize $-\log\left(\mathscr{L}(x)\right)$

     i)

     ii)

     iii) y/a.

More likely, in many realistic problems, is that we would need an iterative procedure to minimize the negative log-likelihood function, i.e. to solve:

$$x_{\mathrm{ML}} = \arg\min_x \; -\log(\mathscr{L}(x))$$

MAXIMUM LIKELIHOOD ESTIMATION

EXAMPLE 2: VECTOR MEASUREMENT CASE IN GAUSSIAN NOISE

- Consider the measurement model: $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}$, where the noise vector $\boldsymbol{n} \in \mathbb{R}^M$ has i.i.d entries, each a zero mean Gaussian random variable, i.e. $\boldsymbol{n} \sim \mathcal{N}(0, \sigma^2)$, $\boldsymbol{y} \in \mathbb{R}^M$ is the measurement vector (measured image), the unknown object is $\boldsymbol{x} \in \mathbb{R}^N$, and $\boldsymbol{A} \in \mathbb{R}^{M\times N}$ is the model for the imaging system.

Solution steps:

1. Write the noise density function: $p_{\boldsymbol{n}}(\boldsymbol{n}) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\|\boldsymbol{n}\|^2}{2\sigma^2}}$
2. Derive likelihood function: $\mathcal{L}(\boldsymbol{x}) = p_n(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|^2}{2\sigma^2}}$
3. Minimize negative log-likelihood, i.e. minimize $-\log\left(\mathcal{L}(\boldsymbol{x})\right)$
   i) What is $\log(\mathcal{L}(x))$?

   $$\log(\mathcal{L}(\boldsymbol{x})) = \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|^2}{2\sigma^2}}\right) = -\log\left(\sqrt{2\pi\sigma^2}\right) + \log\left(e^{-\frac{\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|^2}{2\sigma^2}}\right) = -1/2\log(2\pi\sigma^2) - 1/(2\sigma^2)\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|^2$$

   ii) The minimization problem: $\boldsymbol{x}_{\mathrm{ML}} = \arg\min_{\boldsymbol{x}} \dfrac{1}{2\sigma^2}\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|^2$

   Problem can still be solved, in principle, by differentiating the objective function:

   $$-\frac{\partial}{\partial \boldsymbol{x}}\log(\mathcal{L}(\boldsymbol{x})) = -\boldsymbol{A}^\top(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}))/\sigma^2 = 0$$

   iii) Solve the resulting equation: $-\boldsymbol{A}^\top\left(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}_{\mathrm{ML}}\right)/\sigma^2 = 0 \quad\Rightarrow\quad \boldsymbol{A}^\top\boldsymbol{y} - \boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{x}_{\mathrm{ML}} = 0.$

MAXIMUM LIKELIHOOD ESTIMATION

EXAMPLE 2: VECTOR MEASUREMENT CASE IN GAUSSIAN NOISE

- Consider the measurement model: $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}$, where the noise vector $\boldsymbol{n} \in \mathbb{R}^M$ has i.i.d entries, each a zero mean Gaussian random variable, i.e. $\boldsymbol{n} \sim \mathcal{N}(0, \sigma^2)$, $\boldsymbol{y} \in \mathbb{R}^M$ is the measurement vector (measured image), the unknown object is $\boldsymbol{x} \in \mathbb{R}^N$, and $\boldsymbol{A} \in \mathbb{R}^{M\times N}$ is the model for the imaging system.

Solution steps:

1. Write the noise density function: $p_{\boldsymbol{n}}(\boldsymbol{n}) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\|\boldsymbol{n}\|^2}{2\sigma^2}}$
2. Derive likelihood function: $\mathscr{L}(\boldsymbol{x}) = p_n(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|^2}{2\sigma^2}}$
3. Minimize negative log-likelihood, i.e. minimize $-\log\left(\mathscr{L}(\boldsymbol{x})\right)$

   i) What is $\log(\mathscr{L}(x))$?

   $$\log(\mathscr{L}(\boldsymbol{x})) = \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|^2}{2\sigma^2}}\right) = -\log\left(\sqrt{2\pi\sigma^2}\right) + \log\left(e^{-\frac{\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|^2}{2\sigma^2}}\right) = -1/2\log(2\pi\sigma^2) - 1/(2\sigma^2)\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|^2$$

   ii) The minimization problem: $\boldsymbol{x}_{\mathrm{ML}} = \arg\min_{\boldsymbol{x}} \dfrac{1}{2\sigma^2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2$

   Problem can still be solved, in principle, by differentiating the objective function:

   $$-\frac{\partial}{\partial \boldsymbol{x}} \log(\mathscr{L}(\boldsymbol{x})) = -\boldsymbol{A}^\top(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}))/\sigma^2 = 0$$

   iii) Solve the resulting equation: $-\boldsymbol{A}^\top\left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}_{\mathrm{ML}}\right)/\sigma^2 = 0 \;\Rightarrow\; \boxed{\boldsymbol{A}^\top\boldsymbol{y} - \boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{x}_{\mathrm{ML}} = 0.}$

We already have a plethora of techniques to solve linear systems $\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{A}}\boldsymbol{x}_{\mathrm{ML}}$:

- **Inverse**
- **Iterative methods**

MAXIMUM LIKELIHOOD ESTIMATION

EXAMPLE 3: VECTOR MEASUREMENT CASE IN GAUSSIAN NOISE 2

- Note that when the noise distribution is different, the ML estimate may also be different.

- For instance, if the distribution is Gaussian but the variances for each entry of the noise vector are different from each other

$$p_{\boldsymbol{n}}(\boldsymbol{n}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\boldsymbol{n}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{n}^2}{2}},$$

where $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_N^2 \end{bmatrix}$ is a diagonal matrix with entries.

- In this case, the minimization of the negative log-likelihood function gives:

$$\boxed{\boldsymbol{A}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{A} \boldsymbol{x}_{\mathrm{ML}} = \boldsymbol{A}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{y}}$$

$$P(X \mid Y) = \frac{P(Y \mid X)\, P(X)}{P(Y)}$$

THE UNKNOWN OBJECT TO BE ESTIMATED IS ALSO A REALIZATION OF A RANDOM PROCESS

BAYESIAN METHODS

BAYESIAN METHODS

- The key distinguishing property of **Bayesian methods** is that the **unknown object** is assumed to be a realization of a **random** variable
  - A probability density/distribution can be used to encode any a priori information we we have about the unknown object
- Noise is still assumed to be a realization of a random process
- These can be leveraged for solving the inverse problem
- Need Bayes' rule

MAXIMUM A POSTERIORI (MAP) ESTIMATION

HOW DOES IT WORK?

- Start with our usual discrete model of image formation:

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n},$$

- where:
  - $\boldsymbol{n} \in \mathbb{R}^M$ is unknown random noise process,
  - $\boldsymbol{y} \in \mathbb{R}^M$ is the measurement (measured image/signal)
  - $\boldsymbol{x} \in \mathbb{R}^N$ is the discretization of the unknown object (random variable)
  - $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ is the discrete forward model (for the imaging system)
- The the object $\boldsymbol{x} \in \mathbb{R}^N$ and noise term $\boldsymbol{n} \in \mathbb{R}^M$ are random, as such $\boldsymbol{y} \in \mathbb{R}^M$ is also a random vector.
- A **random vector** simply means that each entry of the 1D array is random:
  - We will assume that they are **independent** but have the same distribution – **independent and identically distributed (or i.i.d)**

MAXIMUM A POSTERIORI (MAP) ESTIMATION

HOW DOES IT WORK?

- Discrete model of image formation:

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n},$$

- Assume noise process $\boldsymbol{n} \in \mathbb{R}^M$ has zero mean
- The distribution of $\boldsymbol{n} \in \mathbb{R}^M$ is assumed to be known, which means the conditional distribution of $\boldsymbol{y}$, denoted as $p_{\boldsymbol{y}|\boldsymbol{x}}\left(\boldsymbol{y}\,|\,\boldsymbol{x}\right) = p_{\boldsymbol{n}}\left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\right)$
- Assume a ***prior*** distribution for $\boldsymbol{x} \in \mathbb{R}^N$ is known $p_{\boldsymbol{x}}(\boldsymbol{x})$
- Then one can find a **conditional density** for $\boldsymbol{x}$ given the measurement $\boldsymbol{y}$:

$$p_{\boldsymbol{x}|\boldsymbol{y}}\left(\boldsymbol{x}\,|\,\boldsymbol{y}\right) = \frac{p_{\boldsymbol{y}|\boldsymbol{x}}\left(\boldsymbol{y}\,|\,\boldsymbol{x}\right)\, p_{\boldsymbol{x}}(\boldsymbol{x})}{p_{\boldsymbol{y}}\left(\boldsymbol{y}\right)} = \frac{p_{\boldsymbol{n}}\left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\right)\, p_{\boldsymbol{x}}(\boldsymbol{x})}{p_{\boldsymbol{y}}\left(\boldsymbol{y}\right)}$$

  - The function $p_{\boldsymbol{x}|\boldsymbol{y}}\left(\boldsymbol{x}\,|\,\boldsymbol{y}\right)$ is called the ***posterior*** *(from 'a posteriori')*
  - The function $p_{\boldsymbol{x}}(\boldsymbol{x})$ is called the ***prior*** *(from 'a priori')*
- *The goal is to find the $\boldsymbol{x}$ the maximizes the posterior distribution – this is called the* ***Maximum a posteriori (MAP) estimate*** $\boldsymbol{x}_{\mathrm{MAP}}$.

MAXIMUM LIKELIHOOD ESTIMATION
HOW DOES IT WORK?

- Discrete model of image formation:
$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n},$$
- The distribution of $\boldsymbol{n} \in \mathbb{R}^M$ is assumed to be known, which means the conditional distribution of $\boldsymbol{y}$, denoted as $p_{y|x}\left(\boldsymbol{y} \mid \boldsymbol{x}\right) = p_n\left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\right)$
- Write the ***prior*** distribution, $p_x(\boldsymbol{x})$, for $\boldsymbol{x} \in \mathbb{R}^N$
- Derive the ***posterior distribution*** for $\boldsymbol{x}$ given the measurement $\boldsymbol{y}$:
$$p_{x|y}\left(\boldsymbol{x} \mid \boldsymbol{y}\right) = \frac{p_n\left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\right) p_x(\boldsymbol{x})}{p_y\left(\boldsymbol{y}\right)}$$
- Maximize (or minimize negative logarithm of) the posterior distribution, i.e.:
$$\boldsymbol{x}_{\mathrm{MAP}} = \arg\min_{\boldsymbol{x}} \; -\log\left(p_n\left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\right)\right) - \log\left(p_x(\boldsymbol{x})\right) + \log\left(p_y\left(\boldsymbol{y}\right)\right)$$

- Note that we can simply ignore $\log\left(p_y\left(\boldsymbol{y}\right)\right)$ because the minimization is over $\boldsymbol{x}$:
$$\boxed{\boldsymbol{x}_{\mathrm{MAP}} = \arg\min_{\boldsymbol{x}} \; -\log\left(p_n\left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\right)\right) - \log\left(p_x(\boldsymbol{x})\right)}$$

MAXIMUM LIKELIHOOD ESTIMATION

EXAMPLE 1: VECTOR MEASUREMENT CASE IN GAUSSIAN NOISE & GAUSSIAN PRIOR

- Consider the measurement model: $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}$, where the noise vector $\boldsymbol{n} \in \mathbb{R}^M$ has i.i.d entries, each a zero mean Gaussian random variable, i.e. $\boldsymbol{n} \sim \mathcal{N}(0, \Sigma_n^2)$, $\boldsymbol{y} \in \mathbb{R}^M$ is the measurement vector (measured image), the unknown object is $\boldsymbol{x} \in \mathbb{R}^N$, and $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ is the model for the imaging system. Assume a Gaussian prior distribution for $\boldsymbol{x} \in \mathbb{R}^N$, i.e $\boldsymbol{x} \sim \mathcal{N}(0, \Sigma_x^2)$
- $\boldsymbol{\Sigma}_x$ and $\boldsymbol{\Sigma}_n$ are covariance matrices (they are symmetric and positive definite).

MAXIMUM LIKELIHOOD ESTIMATION

EXAMPLE 1: VECTOR MEASUREMENT CASE IN GAUSSIAN NOISE & GAUSSIAN PRIOR

Solution steps:

1. Write the noise density function: $p_{\boldsymbol{n}}(\boldsymbol{n}) = \dfrac{1}{\sqrt{2\pi\boldsymbol{\Sigma}_n}} e^{-\frac{\boldsymbol{n}^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{n}}{2}}$
2. Derive conditional density: $p_{\boldsymbol{y}|\boldsymbol{x}}(\boldsymbol{y}\,|\,\boldsymbol{x}) = p_n(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}) = \dfrac{1}{\sqrt{2\pi\boldsymbol{\Sigma}_n}} e^{-\frac{(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})^\top \boldsymbol{\Sigma}_n^{-1}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})}{2}}$
3. Write down the prior distribution: $p_{\boldsymbol{x}}(\boldsymbol{x}) = \dfrac{1}{\sqrt{2\pi\boldsymbol{\Sigma}_x}} e^{-\frac{\boldsymbol{x}^\top \boldsymbol{\Sigma}_x^{-1} \boldsymbol{x}}{2}}$
4. Derive the posterior distribution:

$$p_{\boldsymbol{x}|\boldsymbol{y}}(\boldsymbol{x}\,|\,\boldsymbol{y}) = \frac{1}{p_{\boldsymbol{y}}(\boldsymbol{y})} \frac{1}{\sqrt{2\pi\boldsymbol{\Sigma}_n}} e^{-\frac{(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})^\top \boldsymbol{\Sigma}_n^{-1}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})}{2}} \frac{1}{\sqrt{2\pi\boldsymbol{\Sigma}_x}} e^{-\frac{\boldsymbol{x}^\top \boldsymbol{\Sigma}_x^{-1} \boldsymbol{x}}{2}}$$

   i) Taking its negative natural logarithm?

$$-\log(p_{\boldsymbol{x}|\boldsymbol{y}}(\boldsymbol{x}\,|\,\boldsymbol{x})) = \frac{1}{2}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})^\top \boldsymbol{\Sigma}_n^{-1}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}) + \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{\Sigma}_x^{-1}\boldsymbol{x}$$

   ii) The minimization problem:

$$\boldsymbol{x}_{\mathrm{MAP}} = \arg\min_{\boldsymbol{x}} \frac{1}{2}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})^\top \boldsymbol{\Sigma}_n^{-1}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}) + \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{\Sigma}_x^{-1}\boldsymbol{x}$$

MAXIMUM LIKELIHOOD ESTIMATION

EXAMPLE 1: VECTOR MEASUREMENT CASE IN GAUSSIAN NOISE & GAUSSIAN PRIOR

Solution steps:

1. Derive the posterior distribution:

$$p_{x|y}(\boldsymbol{x}\,|\,\boldsymbol{y}) = \frac{1}{p_y(\boldsymbol{y})}\frac{1}{\sqrt{2\pi\boldsymbol{\Sigma}_n}}e^{-\frac{(\boldsymbol{y}-A\boldsymbol{x})^\top\boldsymbol{\Sigma}_n^{-1}(\boldsymbol{y}-A\boldsymbol{x})}{2}}\frac{1}{\sqrt{2\pi\boldsymbol{\Sigma}_x}}e^{-\frac{\boldsymbol{x}^\top\boldsymbol{\Sigma}_x^{-1}\boldsymbol{x}}{2}}$$

   i) Taking its negative natural logarithm?

$$-\log(p_{x|y}(\boldsymbol{x}\,|\,\boldsymbol{x})) = \frac{1}{2}(\boldsymbol{y}-A\boldsymbol{x})^\top\boldsymbol{\Sigma}_n^{-1}(\boldsymbol{y}-A\boldsymbol{x}) + \frac{1}{2}\boldsymbol{x}^\top\boldsymbol{\Sigma}_x^{-1}\boldsymbol{x}$$

   ii) The minimization problem:

$$\boldsymbol{x}_{\text{MAP}} = \arg\min_x \frac{1}{2}(\boldsymbol{y}-A\boldsymbol{x})^\top\boldsymbol{\Sigma}_n^{-1}(\boldsymbol{y}-A\boldsymbol{x}) + \frac{1}{2}\boldsymbol{x}^\top\boldsymbol{\Sigma}_x^{-1}\boldsymbol{x}$$

- The minimization can be solved, in principle, by differentiating:

$$\boldsymbol{x}_{\text{MAP}} = \left(A^\top\boldsymbol{\Sigma}_n^{-1}A + \boldsymbol{\Sigma}_x^{-1}\right)^{-1}A^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{y}$$

- (Again if the distributions are different, the MAP estimate will also be different.)

MAXIMUM LIKELIHOOD ESTIMATION

EXAMPLE 1: VECTOR MEASUREMENT CASE IN GAUSSIAN NOISE & GAUSSIAN PRIOR

- Consider the **MAP estimate**

$$\boldsymbol{x}_{\text{MAP}} = \left(\boldsymbol{A}^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{A} + \boldsymbol{\Sigma}_x^{-1}\right)^{-1} \boldsymbol{A}^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{y}$$

- When the noise elements have equal variance is equal, i.e. $\boldsymbol{\Sigma}_n = \sigma_n^2 \boldsymbol{I}$ is diagonal, and
- The prior distribution has equal variance $\boldsymbol{\Sigma}_x = \sigma_x^2 \boldsymbol{I}$, then the MAP estimate becomes:

$$\boldsymbol{x}_{\text{MAP}} = \left(\boldsymbol{A}^\top \boldsymbol{A} + \frac{\sigma_n^2}{\sigma_x^2} \boldsymbol{I}\right)^{-1} \boldsymbol{A}^\top \boldsymbol{y}$$

- Notice that this coincides with the **Tikhonov regularized solution** with $\lambda = \sigma_n^2 / \sigma_x^2$.

WHAT WE COVERED TODAY

- Statistical estimation methods
  - Maximum likelihood estimation
  - Maximum a posteriori estimation

TILL NEXT TIME

LINEAR SHIFT VARYING (LSV) IMAGING SYSTEMS