Problem Session 6

Feb. 12\textsuperscript{th} 2016

Orly Liba
Topics

• Photo-electron shot-noise
• SNR calculations
• Deconvolution of an image with Poisson noise
  • Wiener deconvolution
  • Richardson-Lucy
  • Richardson-Lucy + TV prior
• Simulation of a single pixel camera
  • Least norm
  • ADMM + TV prior
Photo-electron shot-noise

• Shot noise occurs during:
  1. Photon counting in optical devices, and associated with the particle nature of light.
  2. The conversion of photons to electrons

• There is statistical variability (=noise) when counting photo-electrons during a specific amount of time.

• Can be modelled by **Poisson statistics**.

• Special property of Poisson statistics:

  **The variance equals the average number of events** \((N)\)

  \(\Rightarrow\) Signal to noise (SNR) \(\frac{N}{\sqrt{N}} = \sqrt{N}\)
Task 1: SNR calculations

You’re asked to compare the SNR of two cameras:
  • Consumer camera
  • sCMOS

At different acquisition modes:
  • Flutter shutter
  • Burst

Calculate the SNR and discuss your results.
SNR calculations

**Flutter shutter**: temporally modulated aperture pattern. Used, for example, for better motion deblurring (see “Coded Exposure Photography: Motion Deblurring using Fluttered Shutter”, Raskar et al.). The result is a single image.

**Burst**: acquires multiple short-exposure images.
SNR calculations

**Consumer camera:** Lots of photons, moderate additive Gaussian noise.

**CMOS microscope sensor:** photon-limited imaging, only a few photons. The sensor is cooled, so no Gaussian noise, but there is photo-electron shot-noise, which has a Poisson distribution.
SNR calculations

SNR = \frac{\text{mean number of photons}}{\text{standard deviation of noise}} = \frac{\mu}{\sigma}

• Adding $m$ signals of strength $k \Rightarrow mk$

• Adding independent Gaussian distributions
  \quad G(\mu_1, \sigma_1^2) + G(\mu_2, \sigma_2^2) \sim G(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)

• Adding independent Poisson distributions
  \quad \text{Pois}(\lambda_1) + \text{Pois}(\lambda_2) \sim \text{Pois}(\lambda_1 + \lambda_2)

Variance = (\text{standard deviation})^2

Both mean and variance
SNR calculations

For each of the four cases, calculate:

• The average number of photons (the signal)
• The standard deviation of the noise
• and divide them

Don’t forget to describe your results and conclusions.
Task 2: Poisson deconvolution

You’re provided with the original image of a cell and code for blurring it and adding Poisson noise.

Your task is to sharpen (deconvolve) the image using:

1. Wiener filtering
2. Richardson-Lucy
3. Richardson-Lucy + a TV prior (with several regularization factors $\lambda$)

Show the resulting image, the convergence of the residual and the MSE and calculate the PSNR.
Poisson deconvolution

Note:

Deconvolution as we saw in week 3 (using ADMM) will not work well. This is because the minimization objective (L2 norm) is optimal for Gaussian noise, not Poisson noise. See notes!

If you want to try ADMM for the Poisson case, complete the bonus question (worth 25 points!).
Poisson deconvolution

Mean squared error (MSE) and the peak signal-to-noise ratio (PSNR):

\[
MSE = \frac{1}{3mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{c=1}^{3} [I_{original}(i,j) - I_{restored}(i,j)]^2
\]

\[
PSNR = 10\log_{10} \left( \frac{\max(I_{original})^2}{MSE} \right)
\]
Poisson deconvolution

Wiener deconvolution:
• Implement in the Fourier domain

\[
H' = \frac{1}{H} \cdot \frac{|H|^2}{(|H|^2 + 1/\text{SNR})}
\]

\[
\text{SNR} = \frac{\bar{I}}{\sigma_{\text{noise}}}
\]

• What is \(\sigma_{\text{noise}}\) for an image with Poisson noise? Consider the properties of the Poisson noise distribution...
Wiener deconvolution

Original

With blurring and noise

Wiener deconvolution

Still blurry...
Poisson deconvolution

Maximum likelihood

We want to find $x$ (the original image) such that it maximizes the probability of $b$ (our noisy & blurry image) given $x$, knowing the blur kernel and assuming Poisson noise.

\[ b \sim \mathcal{P}(Ax) \]
Poisson deconvolution

Maximum likelihood
The probability of a measurement (pixel) \( i \) is given by

\[
p(b_i|x) = \frac{(Ax)_i^b_i e^{-(Ax)_i}}{b_i!}
\]

\( Ax \) is how the image appears on the sensor
→ It determines the average number of photons accumulated in the sensor

Poisson PMF

\[
P(k \text{ events in interval}) = \frac{\lambda^k e^{-\lambda}}{k!}
\]

where
- \( \lambda \) is the average number of events per interval
- \( e \) is the number 2.71828... (Euler’s number) the base of the natural logs
- \( k \) takes values 0, 1, 2, ...
- \( k! \) is the factorial of \( k = k \ast (k-1) \ast (k-2) \ast \ldots \ast 2 \ast 1 \)

This equation is the probability mass function (PMF) for a Poisson distribution.
Poisson deconvolution

**Maximum likelihood**

The joint probability of all measurements (pixels) is expressed by

\[
\begin{align*}
    p(b|x) &= \prod_{i=1}^{M} p(b_i|x) = \prod_{i=1}^{M} \frac{e^{\log((Ax)_i)b_i} e^{-(Ax)_i}}{b_i!} \\
    L(x) &= \log(p(b|x)) = \log(Ax)^T b - (Ax)^T \mathbf{1} - \sum_{i=1}^{M} \log(b_i !)
\end{align*}
\]

and the log likelihood is:

We want to **maximize** \( L(x) \)!
Poisson deconvolution

The Richardson – Lucy algorithms

Assume we have some initial guess of our unknown image \( (x) \), we can apply an iterative correction to the image which will be converged at iteration \( q \) if the estimate does not change after further iterations, i.e.

\[
\frac{x^{(q+1)}}{x^{(q)}} = 1.
\]

At this estimate, the gradient of the objective \( (L(x)) \) will also be zero. The gradient is calculated using matrix calculus.

\[
\nabla L(x) = A^T \text{diag}(Ax)^{-1} b - A^T 1 = 0
\]

Places the elements of \( Ax \) on the diagonal

Sum over the rows of \( A \) (multiplication with ones)
Poisson deconvolution

Richardson – Lucy

\[
\frac{x^{(q+1)}}{x^{(q)}} = 1
\]

\[
\nabla L(x) = A^T \text{diag}(Ax)^{-1} b - A^T 1 = 0 \Rightarrow \text{diag}(A^T 1)^{-1} A^T \text{diag}(Ax)^{-1} b = 1
\]

\[
x^{(q+1)} = 1 \cdot x^{(q)} = (\text{diag}(A^T 1)^{-1} A^T \text{diag}(Ax)^{-1} b) \cdot x^{(q)}
\]

For any initial guess that is purely positive \((x^{(0)} > 0)\), the following iterations will also remain positive. Non-negativity of the estimated image is an important requirement for most imaging applications.
Poisson deconvolution

Richardson-Lucy

Very noisy!!
What should we do?

Add a regularization term (prior)!
Poisson deconvolution

Richardson – Lucy with TV prior

We add the TV term \( (\lambda \| Dx \|_1) \) to \( L(x) \).

After derivation (see notes) we get this update equation:

\[
    x^{(q+1)} = \frac{A^T \left( \frac{b}{Ax} \right)}{A^T 1 - \lambda \left( \frac{D_{xx}}{|D_{xx}|} + \frac{D_{yx}}{|D_{yx}|} \right)} \cdot x^{(q)}
\]
Richardson-Lucy deconvolution

Original

Richardson-Lucy

Richardson-Lucy + TV prior
Poisson deconvolution

Target Image

Blurry and (Poisson) Noisy Image

Wiener Filtering, PSNR: 25.0107

ADMM Reconstruction (no poisson term), PSNR: 24.9851

RL Reconstruction, PSNR: 16.5974

RL+TV Reconstruction, PSNR: 28.6557

Not part of the HW
Poisson deconvolution

The residual is the log likelihood $L(x)$

For RL:

$$
\log (Ax)^T b - (Ax)^T 1 - \sum_{i=1}^{M} \log (b_i!)
$$

For RL + TV:

$$
\log (Ax)^T b - (Ax)^T 1 - \sum_{i=1}^{M} \log (b_i!) - \lambda \|Dx\|_1
$$
Poisson deconvolution

The log likelihood should always increase.

In RL the noise is amplified so much that the image is degraded compared to the beginning.

RL+TV is much better!

The MSE should decrease.
Poisson deconvolution

For part c, show the result, PSNR, log residual and MSE for several values of $\lambda$
Poisson deconvolution

Hints:
• You are given a template which constructs the image. Read it and understand it.
• Start with 100 iterations for RL and RL+TV
• Perform the solution per channel (for loop on the colors)
• Initial guess for x can be anything positive (try uniform or random [0,1]: `rand(imageSize)`)
• Don’t use matrix multiplications! $Ax$ and $A^T x$ should be implemented as functions in Fourier domain.
• Implement point-wise division where needed.
• If you get NaNs, make them zero (a(isnan(a))=0)
• If you get negative values, make them zero (a(a<0)=0)
• We Provide a function for calculating $D_x$, $D_y$ (`opDx.m`)
Task 3: Single pixel camera

Instead of representing an image using $N$ pixels, we record fewer ($M$) pixels by:

1. Multiplying the scene ($x$) with $M$ random masks ($A$)
2. Summing over the all of the resulting pixels
   \[ \rightarrow \text{Each mask results in a single pixel} \]
3. Reconstructing the scene
   a) Least norm solution, using the conjugate gradients method
   b) ADMM with TV prior
Single pixel camera

Formulating the problem:

Image formation for a noisy single-pixel camera:

\[ b = Ax + \eta, \]

where \( \eta \) is an additive, signal-independent Gaussian noise term with standard deviation \( \sigma \), \( b \in \mathbb{R}^M \), \( x \in \mathbb{R}^N \), and \( A \in \mathbb{R}^{M \times N} \).

For compression factors \( \frac{N}{M} = 1, 2, 4, \) and \( 8 \) \( (M < N) \)

We want to reconstruct \( x \).
Single pixel camera

Formulating the problem:

\[ b = Ax + \eta, \]
\[ A \in \mathbb{R}^{M \times N}, \quad M < N \]

→ More variables \((N)\) than equations \((M)\)
→ The system is undetermined, and there are infinite solutions
→ A prior will help us obtain a reasonable solution
  1. Least norm solution
  2. TV prior
Single pixel camera: Least norm solution

Use a prior that is the $\ell_2$-norm on the image: $\|x\|_2^2$

The reconstruction problem is

$$\min_{\{x\}} \frac{1}{2} \|Ax - b\|_2^2 + \|x\|_2^2$$

The solution is obtained by derivation, and the least norm solution is

$$x_{ln} = A^T (AA^T)^{-1} b$$
The conjugate gradient method

• An algorithm for the numerical solution of particular systems of linear equations, namely those whose matrix is symmetric and positive-definite: $Cx = b, \ C^T = C, \ x^T C x > 0$ for all $x \neq 0 \rightarrow x = C^{-1}b$

• Used to calculate $(AA^T)^{-1}b$

• Useful for large matrices, when we want quick convergence

• Use the Matlab function pcg (Preconditioned Conjugate Gradients Method)

• Matlab’s pcg function also supports matrix-free operations, meaning, you can supply a function handle that computes $Cx$ without forming the matrix $C$
Single pixel camera: Least norm solution

Results are not great...
Single pixel camera: ADMM + TV prior

Use a prior that is sparse gradients and minimize the TV norm

\[
\min_{\{x\}} \frac{1}{2} \|Ax - b\|_2^2 + \lambda \|Dx\|_1
\]

regularization

\[
\Gamma(x) = \lambda \|Dx\|_1, \text{ with } D = \begin{bmatrix} D_x^T & D_y^T \end{bmatrix}^T. \quad D \in \mathbb{R}^{2N \times N}
\]

Steps are similar to HW3, only the update step for \( x \) is different
Single pixel camera: ADMM + TV prior

The update for $x$ is:

$(v = z - u)$

$$\text{prox}_{f,\rho}(v) = \left( \frac{A^T A + \rho D^T D}{\tilde{A}} \right)^{-1} \left( \begin{array}{c} A^T b + \rho D^T v \\ \tilde{b} \end{array} \right)$$

See notes for derivation!

$x$ is the solution for $\tilde{A}x = \tilde{b}$

and we can use the conjugate gradients method to solve it.
Single pixel camera: ADMM + TV prior

Compression factor = 1

Compression factor = 2

Experiment with other values of $\lambda$, $\rho$
Single pixel camera: ADMM + TV prior

Compression factor = 4, $\lambda = 1$

Compression factor = 4, $\lambda = 10$

Experiment with other values of $\lambda$, $\rho$
Single pixel camera: ADMM + TV prior

Compression factor = 8, $\lambda = 1$

Compression factor = 8, $\lambda = 10$

Experiment with other values of $\lambda, \rho$
Single pixel camera: ADMM + TV prior

Section c:
- Apply ADMM + TV prior to the provided 12 images
- Use $\lambda=1$, $\rho=10$
- Present your results as a composite image (see next slide)
- Also present a table with the final PSNR values
- Discuss the results
Single pixel camera: ADMM + TV prior

Compression factor

PSNR table

<table>
<thead>
<tr>
<th>Compression factor</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>34.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>27.98</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>24.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>21.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Hints for task 3

• Images are grayscale. 3rd dimension can be used for the N masks.
• Create N random masks (use rand), where N=(# of pixels in image)/(compression)
• Define functions for A, A^T, and AA^T:
  • A performs element-wise multiplication of the 2D input x (which has N elements) with M 2D masks, and sums the results. The output is a vector size M. (b = Ax)
  • A^T multiplies the M masks with the M values of the input and sums the results. The output is a 2D matrix with N elements. (A^Tb)
  • AA^T can be defined using the previous functions. What is the output? (AA^T b will be the size of x)
• Noise is added after applying the masks: \( b = \text{Afun}(I) + \sigma \cdot \text{randn}([N 1]) \);
• For conjugate gradients, use 50 iterations and a tolerance of 10^{-12}
• For ADMM, use the provided functions: opDx, opDtx, which work on non-vectorized images.
• You can also define a handle for D^TD: \( \text{opDtDx} = @(x) \text{opDtx(opDx(x))} \);
• What is the function handle for CG?
• The residual is \( \frac{1}{2} \| Ax - b \|_2^2 + \lambda \| Dx \|_1 \) plot the log residual
Have a nice weekend!

And good luck with the homework!