Image deconvolution, denoising and compression

T.E. Gureyev and Ya.I. Nesterets
15.11.2002
CONVOLUTION

\[ D(x, y) = (I \ast P)(x, y) + N(x, y), \]

where \( D(x, y) \) is the experimental data (registered image),
\( I(x, y) \) is the unknown "ideal" image (to be found),
\( P(x, y) \) is the (usually known) convolution kernel,
* denotes 2D convolution, i.e.

\[ (I \ast P)(x, y) = \iint I(x', y')P(x - x', y - y')dx'dy', \]

and \( N(x, y) \) is the noise in the experimental data.

In real-life imaging, \( P(x, y) \) can be the point-spread function of an imaging system; noise \( N(x, y) \) may not be additive.

Poisson noise: \( D(x, y) = I_{av} \sigma_{av}^2 \text{Poisson}\{I(x, y)/(I_{av} \sigma_{av}^2)\} \)
EXAMPLE OF CONVOLUTION

\[
\begin{array}{c}
\text{3\% Poisson noise} \\
\text{10\% Poisson noise}
\end{array}
\]
DECONVOLUTION PROBLEM

\[ D(x, y) = (I \ast P)(x, y) + N(x, y) \quad (*) \]

Deconvolution problem: given \( D, P \) and \( N \), find \( I \) (i.e. compensate for noise and the PSF of the imaging system)

Blind deconvolution: \( P \) is also unknown.

Equation (*) is mathematically ill-posed, i.e. its solution may not exist, may not be unique and may be unstable with respect to small perturbations of "input" data \( D, P \) and \( N \). This is easy to see in the Fourier representation of eq.(*)

\[ \hat{D}(\xi, \eta) = \hat{I}(\xi, \eta) \hat{P}(\xi, \eta) + \hat{N}(\xi, \eta) \quad (#) \]

1) Non-existence: \( \hat{D}(\xi, \eta) \neq 0, \hat{P}(\xi, \eta) = \hat{N}(\xi, \eta) = 0 \Rightarrow \hat{I}(\xi, \eta) = ? \)
2) Non-uniqueness: \( P(\xi, \eta) = 0, \hat{D}(\xi, \eta) = \hat{N}(\xi, \eta) \Rightarrow \hat{I}(\xi, \eta) = \text{any} \)
3) Instability: \( P(\xi, \eta) = \varepsilon \Rightarrow \hat{I}(\xi, \eta) = [\hat{D}(\xi, \eta) - \hat{N}(\xi, \eta)] / \varepsilon \)
A SOLUTION OF THE DECONVOLUTION PROBLEM

Convolution: \[ \hat{D}(\xi,\eta) = \hat{I}(\xi,\eta)\hat{P}(\xi,\eta) + \hat{N}(\xi,\eta) \quad (#) \]

Deconvolution: \[ \hat{I}(\xi,\eta) = [\hat{D}(\xi,\eta) - \hat{N}(\xi,\eta)]/ \hat{P}(\xi,\eta) \quad (!) \]

We assume that \( \hat{P}(\xi,\eta) \neq 0 \) (otherwise there is a genuine loss of information and the problem cannot be solved). Then eq.(!) provides a nice solution at least in the noise-free case (as in reality the noise cannot be subtracted exactly).
NON-LOCALITY OF (DE)CONVOLUTION

\[(I * P)(x, y) = \iint I(x', y')P(x-x', y-y')dx'dy'\]

The value of convolution \(I*P\) at point \((x,y)\) depends on all those values \(I(x',y')\) within the vicinity of \((x,y)\) where \(P(x-x', y-y')\neq 0\). The same is true for deconvolution.

Convolution with a single pixel wide mask at the edges

Deconvolution (the error is due to the non-locality and the 1-pixel wide mask)
EFFECT OF NOISE

In the presence of noise, the ill-posedness of deconvolution leads to artefacts in deconvolved images: \( \hat{I}(\xi, \eta) \equiv \hat{D} / \hat{P} \)

The problem can be alleviated with the help of regularization

\[
\hat{I}(\xi, \eta) = \hat{D}(\xi, \eta) \hat{P}^* (\xi, \eta) / [ | \hat{P}(\xi, \eta) |^2 + \alpha ]
\]

without regularization\(
\leftrightarrow
\)

3% noise in the experimental data\((\ast)^{-1}\)

with regularization\leftrightarrow
EFFECT OF NOISE. II

In the presence of stronger noise, regularization may not be able to deliver satisfactory results, as the loss of high frequency information becomes very significant. Pre-filtering (denoising) before deconvolution can potentially be of much assistance.

10% noise in the experimental data

\[ (*)^{-1} \]

without regularization \(\rightarrow\)

with regularization \(\rightarrow\)
DECONVOLUTION METHODS

Two broad categories

(1) Direct methods
Directly solve the inverse problem (deconvolution).
Advantages: often linear, deterministic, non-iterative and fast. 
Disadvantages: sensitivity to (amplification of) noise, difficulty in incorporating available \textit{a priori} information. 
Examples: Fourier (Wiener) deconvolution, algebraic inversion.

(2) Indirect methods
Perform (parametric) modelling, solve the forward problem (convolution) and minimize a cost function. 
Disadvantages: often non-linear, probabilistic, iterative and slow. 
Advantages: better treatment of noise, easy incorporation of available \textit{a priori} information. 
Examples: least-squares fit, Maximum Entropy, Richardson-Lucy, Pixon
DIRECT DECONVOLUTION METHODS

Fourier (Wiener) deconvolution
Based on the formula
\[ \hat{I}(\xi, \eta) = \hat{D}(\xi, \eta) \hat{P}^*(\xi, \eta) / [ | \hat{P}(\xi, \eta) |^2 + \alpha ] \]
Requires 2 FFTs of the input image (very fast)
Does not perform very well in the presence of noise

Convolution with 10% noise
ITERATIVE WIENER DECONVOLUTION

The method proposed by A.W. Stevenson
Builds deconvolution as

\[ I^{(n)} = W(D, \alpha) + \sum_{k=0}^{n-1} W(I^{(k)} \ast P - D, \alpha_k) \]

Requires 2 FFTs of the input image at each iteration
Can use large (and/or variable) regularization parameter \( \alpha \)

Convolution with 10% noise
Richardson-Lucy (RL) algorithm

If PSF is shift invariant then RL iterative algorithm is written as

\[
I^{(i+1)} = I^{(i)} \operatorname{Corr}(D / (I^{(i)} \ast P ), P)
\]

Correlation of two matrices is \(\operatorname{Corr}(g, h)_{n,m} = \sum_i \sum_j g_{n+i,m+j} h_{i,j}\)

Usually the initial guess is uniform, i.e. \(I^{(0)} = \langle D \rangle\)

Advantages:
1. Easy to implement (no non-linear optimisation is needed)
2. Implicit object size \((I_n^{(0)} = 0 \Rightarrow I_n^{(i)} = 0 \ \forall i)\)
   and positivity constraints \((I_n^{(0)} > 0 \Rightarrow I_n^{(i)} > 0 \ \forall i)\)

Disadvantages:
1. Slow convergence in the absence of noise and instability in the presence of noise
2. Produces edge artifacts and spurious "sources"
No noise

Original Image

1000 RL iterations

50 RL iterations
3% noise

Original Image

Data

6 RL iterations

20 RL iterations
Bayesian methods

Joint probability of two events $p(A, B) = p(A) p(B | A)$

$p(D, I, M) = p(D | I, M) p(I, M) = p(D | I, M) p(I | M) p(M)$

$= p(I | D, M) p(D, M) = p(I | D, M) p(D | M) p(M)$

$I$ – image $\xrightarrow{M - \text{model}}$ $D$ – data

\[ p(I | D, M) = \frac{p(D | I, M) p(I | M)}{p(D | M)} \]  \hspace{1cm} (1)

\[ p(I, M | D) = \frac{p(D | I, M) p(I | M) p(M)}{p(D)} \]  \hspace{1cm} (2)

$p(I | D, M)$ or $p(I, M | D)$ – inference

$p(I, M)$, $p(I | M)$ and $p(M)$ – priors

$p(D | I, M)$ – likelihood function
Goodness-of-fit (GOF) and Maximum Likelihood (ML) methods

Assumes image prior \( p(I \mid M) = \text{const} \) and results in maximization with respect to \( I \) of the likelihood function \( p(D \mid I, M) \)

In the case of Gaussian noise (e.g. instrumentation noise)

\[
p(D \mid I, M) = Z^{-1} \exp(-\chi^2 / 2) \quad \text{(standard chi-square distribution)}
\]

where \( \chi^2 = \sum_k \left( \frac{(I \ast P)_k - D_k}{\sigma_k^2} \right)^2 \), \( Z = \prod_k (2\pi\sigma_k^2)^{1/2} \)

In the case of Poisson noise (count statistics noise)

\[
p(D \mid I, M) = \prod_k \frac{(I \ast P)_k^{D_k}}{D_k} \exp\left(-\left(I \ast P\right)_k\right)
\]

Without regularization this approach typically produces images with spurious features resulting from over-fitting of the noisy data
GOF

Data → No noise

Original Image

Deconvolution → 3% noise
Maximum Entropy (ME) methods  
(S.F.Gull and G.J.Daniell; J.Skilling and R.K.Bryan)  
ME principle states that *a priori* the most likely image is the one which is completely flat  

**Image prior**: \( p(I \mid M) = \exp(\lambda S) , \)  
where \( S = -\sum_{i} p_{i} \log_{2} p_{i} \) is the image entropy, \( p_{i} = I_{i} / \sum_{i} I_{i} \)  

**GOF term** (usually): \( p(D \mid I, M) = Z^{-1} \exp(-\chi^{2} / 2) \)  

**The likelihood function**: \( p(I \mid D, M) \propto \exp(-L + \lambda S) , \quad L = \chi^{2} / 2 \)  

+ tends to suppress spurious sources in the data  
- can cause over-flattenning of the image  
! the relative weight of GOF and entropy terms is crucial
ME deconvolution

\[ p(I \mid D, M) \propto \exp(-L + \lambda S) \]

\[ \lambda = 2 \quad \lambda = 5 \quad \lambda = 10 \]
Pixon method

Image prior is \( p(I \mid M) = p(\{N_i\}, n, N) = \frac{N!}{(n^N \prod N_i!)} \)

where \( N_i \) is the number of units of signal (e.g. counts) in the \( i \)-th element of the image, \( n \) is the total number of elements,

\( N = \sum_i N_i \) is the total signal in the image

Image prior can be maximized by

1. decreasing the total number of cells, \( n \), and
2. making the \( \{N_i\} \) as large as possible.

\[
I(x) = (I_{\text{pseudo}} * K)(x) = \int dy \ K( (x - y) / \delta(x) ) I_{\text{pseudo}}(y)
\]

\( K \) is the pixon shape function normalized to unit volume

\( I_{\text{pseudo}} \) is the “pseudo image”
No noise
3% noise
DOES A METHOD EXIST CAPABLE OF BETTER DECONVOLUTION IN THE PRESENCE OF NOISE ???

1) Test images can be found on "kayak" in "common/DemoImages/DeBlurSamples" directory

2) Some deconvolution routines have been implemented online and can be used with uploaded images. These routines can be found at "www.soft4science.org" in the "Projects… On-line interactive services… Deblurring on-line" area