Chapter 7

Case Study 1 Image Retrieval A: Image deBlurring

Different Types of Image Blur

- Defocus blur
 - --- Depth of field effects
- Scene motion
 - --- Objects in the scene moving
- Camera shake
 - --- User moving hands





$$I_{\text{PSF}}(x, y, z) = \left| \iint P(k_x, k_y) e^{i2\pi (k_x x + k_y y)} e^{i2\pi k_z z} dk_x dk_y \right|^2$$

The defocus phase is described by the $\exp[i2\pi k_z z]$ factor.



Image formation model: Convolution



Blind vs. non-blind deconvolution

Blind vs. non-blind deconvolution

Non-blind





Image deconvolution is ill-posed



Why is this hard?



Need more information !!!!

Case study 1: Deblurring with Blind Deconvolution







Original Image



Probabilistic Model of Image Formation

X = sharp image Y = observed (blurry) image K = blur kernel

$P(K, X | Y) \sim P(Y | K, X) P(X) P(K)$

Posterior

1. Likelihood 2. Image 3. Blur (Reconstruction constraint)

prior

prior

Deconvolution with prior



2. Image prior:



Distribution of gradients

3. Blur prior:



Positive & Sparse

Likelihood P(Y | K, X)



Image prior P(X): Use parametric model (mixture of Gaussians) of sharp image statistics



Blurry images have different statistics



Blur prior P(K): Positive and Sparse



The obvious thing to do: a MAP Solver

X = sharp image Y = observed (blurry) image K = blur kernel

$P(K, X | Y) \sim P(Y | K, X) P(X) P(K)$

Optimize this objective function for K and X

- This is called Maximum a-Posteriori (MAP)



No success!

Deconvolution with prior using Variational Bayesian approach (link to pdf paper)



Deconvolution with prior using Variational Bayesian approach (link to m-scripts)

1. Pre-processing

- 2. Kernel estimation
 - Multi-scale approach

Input image



- 3. Image reconstruction
 - Standard non-blind deconvolution routine

Original code available at:

http://www.inference.phy.cam.ac.uk/jwm1003/train_ensemble.tar.gz

Preprocessing



Initialization



Inferring the kernel: multiscale method



Image Reconstruction



Case Study 1 Image Retrieval: B. Image Deconvolution

Case Study 1 Image Retrieval B: Image Deconvolution

7.2 Framework of Bayes rule for image retrieval applications

To show how to implement "Bayesian Inference" in image processing, specifically image deconvolution, let us consider the following image formation process g = Hf using a shift-invariant optical system *H*. Let's denote the object to be imaged as *f* and *g* as the image. In real applications, noise *N* and background *b* will be encountered during the imag taking process, which results in g = N[Hf + b].

To retrieve *f* from an image *g*, we often implement a least-squareminimization (lsqmin) solver for $\partial \Psi / \delta f = 0$ with $\Psi = \|Hf - g\|_{L^2}^2$. The associated mathematics yields $H^T Hf - H^T g = 0$ and gives a solution $\hat{f} = \frac{H^T g}{H^T H}$. Unfortunately, this is an ill-posed problem because the eigenvalues of $H^T H$ can become extremely small, thus \hat{f} will diverge during iteration.

To solve the problem, we can regularize the data using some prior knowledge as a constraint. If image formation process is linear, the data regularization yields the result of Tikhonov-Miller (or Wiener) filtering. In the far-field imaging, $k_z = \sqrt{(2\pi/\lambda)^2 - k_x^2 - k_y^2}$ is real, implying information is missing outside an angular cone of spatial frequencies. If we can recover those missing information by using nonlinear image formation, a superresolution imaging may be achieved.

For the case of linear image formation, we can obtain some prior knowledge from that the energy involved in an image is constant, *i.e.*, $E = ||f||^2 = \text{constant}$. Thus, the lsqmin problem $\delta L/\delta f = 0$ with a constraint

$$L = \left\| Hf - g \right\|_{L^2}^2 + \gamma \left(\left\| f \right\|_{L^2}^2 - E \right) \text{ has a solution of } \hat{f} = \frac{H^T g}{H^T H + \gamma}, \text{ showing that } \gamma$$

increases, the degree of smoothing in f increases.

To proceed further, we can construct a Maximum a Posteriori (MAP) solver to yield a solution by maximizing the conditional probability of P(f|g). First, we apply the Bayes Rule to rewrite $P(f|g) = \frac{P(g|f)P(f)}{P(g)}$. Here the marginal probability P(g), which is irrespective of our knowledge about f, reflects how reliable of evidence g is.

For image retrieval, the prior knowledge about object P(f) can be expressed as $P(f) = e^{-\frac{1}{2\tau^2} \|C(f-m)\|_{L^2}^2}$ using τ as a penalty to scale the deviation cost of the model *m* (represented by a set of model parameters) from a true object *f*. Here C represents a regularization matrix. An image is usually recorded with a camera, which is contaminated by noise. Assume that the noise source follows a Poisson statistics with a likelihood probability

 $P(g \mid f) = \prod_{i=1}^{K \in \text{ pixels on camera}} \frac{\mu_i^{N_i} e^{-\mu_i}}{N_i!} \text{ in term of } \mu_i = q[Hf + b]_i \text{ (the mean number of photons detected at pixel i) and } N_i = q g_i \text{ (} q = \text{quantum yield and } g_i = \text{the number of photons arriving at pixel } i \text{ in a time interval).}$

Based on the probabilities, we can implement a MAP solver as follows: First define a Lagrangian of the problem as

$$L = -\log P(f \mid g) = \sum_{i \in \text{pixels}} Hf_i - g_i^T \ln(Hf + b)_i + \gamma \|C(f - m)\|_{L^2}^2 \text{ and then invokes a}$$

nonnegative constraint on f by expressing $f = e^2 \ge 0$. By using variational principle $\delta L/\delta f = 0$ with $L = \sum_{i \in \text{pixels}} (He^2)_i - g_i^T \ln(He^2 + b)_i + \gamma' ||e^2||_{L^2}^2$, we then have an equation for the MAP solution. We can solve the equation for a set of model parameters that will maximize the posteriori probability.

Case Study (TV/L2-Deconv): Formulate Image Deconvolution as a Constraint Minimization Problem using L2-Norm as a Distance Metric for **Data Fidelity** and Total Variation as **Data Constraint**

In this case study, we will illustrate how to formulate an image deconvolution as a constraint minimization problem and to devise a solver to retrieve the object function f from a given image data g.

minimize $\left[\frac{\mu}{2} \| Hf - g \|_{L_2}^2 + \| Df \|_{L_1} \right],$

which includes the data fidelity $\frac{\mu}{2} \|Hf - g\|_{L_2}^2$ and data constraint $\|Df\|_{L_1}$.

First, let us define u = Df, the mathematical problem becomes

minimize
$$\left[\frac{\mu}{2} \| Hf - g \|_{L_2}^2 + \| u \|_{L_1} \right]$$
 s.t. $u = Df$.

Next an augmented Lagrangian for the constraint minimization problem will

$$L(f, u, \lambda) = \frac{\mu}{2} \|Hf - g\|_{2}^{2} + \|u\|_{1} - \lambda^{T} (u - Df) + \frac{\rho}{2} \|u - Df\|_{2}^{2}$$

= $\frac{\mu}{2} (Hf - g)^{T} \cdot (Hf - g) + \|u\|_{1} - \lambda^{T} (u - Df) + \frac{\rho}{2} (u - Df)^{T} \cdot (u - Df)$
= $\frac{\mu}{2} (f^{T} H^{T} Hf - f^{T} H^{T} g - g^{T} Hf + g^{T} g) + \|u\|_{1} - \lambda^{T} (u - Df) + \frac{\rho}{2} (u^{T} u - u^{T} Df - f^{T} D^{T} u + f^{T} D^{T} Df)$

Invoking the variational calculus algebra $\delta(f^T A)/\delta f = A$, $\delta(Af)/\delta f = A^T$, and $\delta(f^T A f)/\delta f = 2Af$,

 $\delta L/\delta f = 0$ yields

$$(\mu |H|^2 f + \rho D^T D) = \mu H^T g + \rho D^T u - D^T \lambda , \qquad (A)$$

and $\delta L / \delta u = 0$

$$\lambda = \rho(u - Df) \quad . \tag{B}$$

A state-of-the-art solver (alternating direction method of multipliers, ADMM) can be devised to provide a solution to the constraint minimization problem. ADMM is an algorithm that solves convex optimization problems by breaking them into smaller pieces:

Subproblem 1: $f_{k+1} = \arg\min_{f} \left[\frac{\mu}{2} \|Hf_k - g\|_2^2 - \lambda_k^T (u_k - Df_k) + \frac{\rho}{2} \|u_k - Df_k\|_2^2\right]$

$$\rightarrow f_{k+1} = F^{-1} \{ \frac{F[\frac{\mu}{\rho} H^T g + D^T u - \frac{1}{\rho} D^T \lambda]}{\frac{\mu}{\rho} |F(H)|^2 + (|F(D_x)|^2 + |F(D_y)|^2 + |F(D_z)|^2)} \},$$

Subproblem 2: $u_{k+1} = \arg\min_{u} \left[|u|_1 - \lambda_k^T (u - Df_{k+1}) + \frac{\rho}{2} \|u - Df_{k+1}\|_2^2 \right]$

$$\rightarrow u_{k+1} = \frac{\rho}{2} \left\| (Df - u + \frac{1}{\rho} \lambda) \right\|_{2}^{2},$$

Subproblem 3: $\lambda_{k+1} = \lambda_k - \rho(u_{k+1} - Df_{k+1})$.

Case Study 1 Image Retrieval C: Phase Retrieval

7.3 Phase Information

The random fluctuations of the atmosphere and the mechanical instability of an interferometer prevent the phase information of μ_P at the exit pupil plane (*the normalized version of the mutual intensity function*) to be extracted. Thus, the question is often that *can we deduce the object information from* $|\mu_P|$ *alone*? Let us examine the following figures:



Note if $|\mu_P(\Delta x, \Delta y)|^2$ is given, then

 $I_i(\vec{u}) = FT[|\mu_P(\Delta x, \Delta y)|^2] = \int I_o(\vec{\xi})I_o(\vec{\xi} - \vec{u}) dA_{\vec{\xi}}$, which can be used to obtain the separation of two small objects like



If the object of interest happens to have a point source near it and separated at a proper distance (like a complex galaxy with a nearby point bright star) as depicted below, we can then extract the object information in this case.



7.3.1 Phase Retrieval Problem

If an object is bounded (i.e., nonzero only over a finite domain on the object plane ξ), and assume $I_o(\xi) = 0$ for all $\xi < 0$, then

$$\begin{split} \mu(\Delta x) &= \int_{-\infty}^{\infty} d\xi \, I_o(\xi) e^{j2\pi\xi\cdot\Delta x} = \int_0^{\infty} d\xi \, I_o(\xi) e^{j2\pi\xi\cdot\Delta x} \\ &= \mu_r + j\mu_i \end{split}$$
, which is analytic in the

upper half of the complex z plane, *i.e.*, the real and imaginary parts of $\mu(\Delta x)$ are connected by

$$\mu_r(\Delta x) = \frac{1}{\pi} \int_{-\infty}^{\infty} dz \frac{\mu_i(z)}{z - \Delta x}, \quad \mu_i(\Delta x) = -\frac{1}{\pi} \int_{-\infty}^{\infty} dz \frac{\mu_r(z)}{z - \Delta x}$$

Let $\mu(\Delta x) = \left| \mu(\Delta x) \right| e^{j\alpha(\Delta x)}$, then $\alpha(\Delta x) = -\frac{1}{\pi} \int_{-\infty}^{\infty} dz \frac{\ln \left| \mu(z) \right|}{z - \Delta x}$.

Note $\ln |\mu(z)|$ is not necessary to be analytic in the upper half of the *z*-plane, even if $\mu(z)$ is known to be analytic there. But if

$$\int_{-\infty}^{\infty} d\Delta x \left| \mu(\Delta x) \right|^2 < \infty \quad and \quad \int_{-\infty}^{\infty} d\Delta x \frac{\ln \left| \mu(\Delta x) \right|}{(\Delta x)^2 + 1} < \infty,$$

then

$$\alpha(\Delta x) = -\frac{1}{\pi} \int_{-\infty}^{\infty} dz \, \frac{\ln \left| \mu(z) \right|}{z - \Delta x} + \sum_{\substack{n \in zeroes \text{ of } \mu(z) \\ \text{ in the uhp of } z}} \arg(\frac{\Delta x - z_n}{\Delta x - z_n^*}) \, .$$

Unfortunately, in general there is no good way to know the locations of the zeroes. Ambiguities exist in the general case. Some algorithms may be useful to conquer the difficulty.

■ *Phase retrieval with iteration* (the Fienup scheme)

Assume: 1. $|\mu(\Delta x)|$ has been measured;

2. $I_o(\xi)$ is bounded; and positive $I_o(\xi) \ge 0$.

Under the constraints, a successful iteration procedure to retrieve the phase and the object information had been developed by Fienup, which is sketched as followed:

- 1. From the known $|\mu(\Delta x)|$, first calculate $I_o^{(1)}(\vec{\xi}) = FT^{-1}[|\mu(\Delta \vec{x})|]$;
- 2. By setting the negative values of $I_o^{(1)}(\vec{\xi})$ to be zero, and the nonzero $I_o^{(1)}(\vec{\xi})$ outside the bound of $I_o(\vec{\xi})$ to be zero $\Rightarrow \hat{I}_o^{(1)}(\vec{\xi})$.
- 3. Take $\tilde{\mu}^{(2)}(\Delta \vec{x}) = FT[\hat{I}_o^{(1)}(\vec{\xi})].$
- 4. Calculate $\mu^{(2)}(\Delta \vec{x}) = \left| \mu(\Delta \vec{x}) \right| e^{j \arg[\tilde{\mu}^{(2)}(\Delta \vec{x})]}$.
- 5. Repeat steps 1 to 4.

A typical result of this phase retrieval procedure is summarized below



7.3.2 Phase Retrieval in a Realistic Situation

Iterative Projection Algorithms for Phase Retrieval

By S. Marchesinia, Rev. of Scientific Instrum.78, 011301 (2007)

Iterative projection algorithms can serve as a substitute of lenses in an optical imaging system to recombine light scattered by illuminated objects in a numerical manner. When the intensity pattern scattered by an object is collected by a camera, the phase information is missing. Consider an object of density $\rho(\vec{r})$ with \vec{r} being the coordinates in the object real space. The diffraction pattern generated

$$I(\vec{k}) = \tilde{\rho}^+(\vec{k})\tilde{\rho}(\vec{k}) = |\tilde{\rho}(\vec{k})|^2$$

is equal to the modulus square of the Fourier-transform $\tilde{\rho}(\vec{k})$. The inverse Fouriertransform $IFT[I(\vec{k})]$ of the measured intensity yields the autocorrelation $\rho(-\vec{r}) \otimes \rho(\vec{r})$ of the object. Since the intensity $I(\vec{k})$ represents the FT of the autocorrelation function, and the autocorrelation is twice as large as the object, the diffraction pattern intensity should be sampled at least twice as finely as the amplitude to capture all possible information on the object. It can be shown that less than critical sampling was sufficient to solve the phase problem. This was possible because the number of measured intensities in the 2D and 3D phase retrieval problems is larger than the number of resolution elements in the object.

Coherence is required to properly sample the FT of the autocorrelation of the object. According to the Schell theorem, the autocorrelation of the illuminated object obtained from the recorded intensity is multiplied by the complex degree of coherence. The optical beam needs to fully illuminate the isolated object, and the degree of coherence must be larger than its autocorrelation.

Phase retrieval problem in optics was solved by using the knowledge that the object being imaged is isolated, indicating the solution shall be 0 outside a region called support S

(i.e., $\rho(\vec{r}) = 0$, if $\vec{r} \notin S$, resulting in a modulus constraint of $I(\vec{k}) = |\sum_{\vec{r} \in S} \rho(\vec{r})e^{i\vec{k}\cdot\vec{r}}|^2$). If

the number of independent equations equals the number of unknowns, the system has a single solution. The intersection of these constraints (both the finite support and given modulus) forms the solution. Unfortunately this system of equations is difficult to solve, and has an enormous number of local minima. The presence of noise and limited prior knowledge further loose the constraints and thereby increases the number of solutions within the noise level and constraints.



In the past decades, researchers had successfully developed several Iterative Projection Algorithms to solve the phase retrieval problem. As illustrated in the above figure, these algorithms try to find the intersection between two sets (i.e., S: the finite support and M: the given modulus). The search for the intersection is based on the information obtained by projecting the current estimate on the two sets. When the image belongs to both sets simultaneously, we have reached a solution. An error metric can be used to evaluate the distance between the current estimate and a given set.

We can devise a projection P_s of $\rho(\vec{r})$ onto S by setting $\rho(\vec{r})$ to 0 outside the support S, while leaving the rest of the values unchanged

$$P_{\!s}
ho(ec r) = egin{cases}
ho(ec r) & ext{if } ec r \in S \ 0 & ext{if } ec r
ot s \end{cases}.$$

And the complementary projector of $P_{\overline{s}} = I - P_s$ can also be defined.

In an intensity measurement we obtain the amplitude or modulus in every pixel that defines a circle in a complex plane. These circles define the modulus constraint. When every complex-valued pixel lies on the

circle defined by the corresponding modulus, the image that satisfies this constraint, which is nonconvex, belongs to the modulus set. The projection of a point in each complex plane onto the corresponding circle is accomplished by taking the point on the circle closest to the current one, setting the modulus to the measured one, and leaving the phase unchanged (which is a nonlinear operator),

$$\tilde{P}_m(\vec{k})\tilde{
ho}(\vec{k}) = \tilde{P}_m(\vec{k}) \mid \tilde{
ho}(\vec{k}) \mid e^{i\phi(\vec{k})} = \sqrt{\tilde{I}(\vec{k})} \cdot e^{i\phi(\vec{k})}$$
, where $P_m = F^{-1}\tilde{P}_m(\vec{k})F$.

A projector P can be viewed as an operator that takes to the closest point in a set from the current point. A repetition of the same projection is equal to one projection alone $P^2 = P$, so its eigenvalues must be 0 or 1. Another useful operator for phase retrieval is the reflector defined by R = I + 2(P - I) = 2P - I, which applies the same step as the projector but moves twice as far.



L2-norm of $\rho(\vec{r})$ can be defined as $\|\rho\|^2 = \rho^+ \cdot \rho = \sum_k \frac{|\tilde{\rho}(\vec{k})|^2}{\sigma^2(\vec{k})} / \sum_k (\frac{1}{\sigma^2(\vec{k})})$ and can

be employed to quantify the distance difference of $\rho(\vec{r})$ from a set. $\|P\rho - \rho\|^2$ denotes the distance from the current point $\rho(\vec{r})$ to the set. Thereby, the errors in real and reciprocal space can be defined in terms of their distance to the corresponding sets as $\varepsilon_s(\rho) = \|P_s\rho - \rho\|$, and

$$\varepsilon_m(\rho) = \left\| P_m \rho - \rho \right\|.$$

Projector P_m moves $\rho(\vec{r})$ to the closest minimum of $\varepsilon_m^{2}(\rho)$ as shown by

$$\varepsilon_m^2(P_m\rho) = \left\| P_m P_m \rho - P_m \rho \right\|^2 = \left\| P_m \rho - P_m \rho \right\|^2 = 0$$

This also yields a simple relation with the gradient $abla_{\rho} \varepsilon_{m}^{2}$ as shown below

$$P_m \rho = \rho + (P_m - I)\rho = \rho - \frac{1}{2} \nabla_\rho \varepsilon^2_{\ m}(\rho) = \rho - \varepsilon_m(\rho) \nabla_\rho \varepsilon_m(\rho) \,.$$

Similarly, the projector P_s minimizes the error $\varepsilon_s^{-2}(\rho)$ as

$$(I - P_s)\rho = P_{\overline{S}}\rho = \rho_{\overline{S}} = \frac{1}{2}\nabla_{\rho}\varepsilon^2_{\ s}(\rho) = \varepsilon_s(\rho)\nabla_{\rho}\varepsilon_s(\rho).$$

The PR algorithms require a starting point $\tilde{\rho}^{(0)}(\vec{k}) = \sqrt{I(\vec{k})}$, which can be usefully generated by assigning a random phase to the measured object amplitude modulus in the Fourier domain. The simplest algorithm called error reduction (ER) is iterated via

$$\rho^{(n+1)} = P_s P_m \rho^{(n)} = P_s \rho^{(n)} - \frac{1}{2} P_s \nabla_\rho \varepsilon_m^2(\rho) = P_s \rho^{(n)} - \frac{1}{2} \nabla_s \varepsilon_m^2(\rho)$$

The solvent flipping (SF) algorithm is obtained by replacing the support projector P_s with its reflector

 $R_s=2P_s-I$ as $\,\rho^{(n+1)}=R_sP_m\rho^{(n)}$. The hybrid input-output (HIO) is based on nonlinear feedback control theory and can be expressed as

$$\rho^{(n+1)} = \begin{cases} P_m \rho^{(n)}(\vec{r}) \text{ if } \vec{r} \in S\\ (I - \beta P_m) \rho^{(n)}(\vec{r}) \text{ otherwise} \end{cases}$$



As illustrated in the above figure, ER simply projects back and forth between these two sets, and moves along the support line in the direction of the intersection. SF projects onto the modulus, reflects on the support, and moves along the reflection of the modulus constraint onto the support. The solvent flipping algorithm is slightly faster than ER thanks to the increased in the angle between projections and reflections.

More advanced algorithms can also be devised to accelerate and improve the convergence to the global minimum. For examples, the difference map (DM), the averaged successive reflections (ASR), and the Hybrid Projection Reflection (HPR) had been successfully implemented for optical PR problem. HIO and variants ASR, DM, HPR move in a spiral around the intersection, eventually reaching the intersection. For similar β , RAAR behaves somewhere in between ER and HIO with a sharper spiral, reaching the solution much earlier.



The basic features of the iterative projection algorithms can be understood by a simple model of two lines intersecting (a). The aim is to find the intersection. The ER algorithm and the solvent flipping algorithms

converge in some gradient-type fashion (the distance to the two sets never increases), the solvent flip method being slightly faster when the angle between the two lines is small. HIO and variants move following a spiral path.



The above figure depicts the error metric $\varepsilon_m(\rho)$ in a simple two-dimensional phaseretrieval problem. The behavior of the ER algorithm toward the local minima is presented in (a). The presence of local minima will cause stagnation of steepest and conjugate gradient methods, preventing global convergence as shown by (c). The ability to escape local minima demonstrated by input-output feedback-based algorithms (see (d)) makes them superior to the methods based on simple gradient minimization of the error. However, as in the ER algorithm, the step length is not optimized, the algorithm keeps moving in the same direction for several steps, and sometimes overshoots. Combining the ideas of the conjugate gradient or the steepest descent methods and IO feedback could considerably speed-up convergence. Optimization of the step length by increasing a multiplication factor until the current and next search directions becomes perpendicular to one another (e). In analogy to the conjugate gradient method, one could substitute the search direction $\Delta \rho$ following the conjugate gradient scheme (see f). A more robust strategy involves replacing the one-dimensional search with a 2D optimization of the saddle point (SO2D, see g). Solving the 2D min–max problem following the conjugate gradient scheme yields the best performance (see i).

Algorithm	No. of iterations for 50% success	Success after 10 000 iterations
HIO/HPR	2790	82%
HIO/HPR+ER	2379	82.6%
ASR	1697 ^a	42%
SO2D	656	100%
SO4D	605	100%
Others	>10 000	0%

In short, **HIO appears to be the most effective algorithm**, and it is significantly improved in terms of speed and reliability **when the 2D step size optimization is applied**.

 $\overline{{}^{a}42\%}$ success, the algorithm either reconstruct in a limited number of iterations or never.