A spatially adaptive nonparametric regression image deblurring

Vladimir Katkovnik, Karen Egiazarian,
Alessandro Foi, Jaakko Astola

Signal Processing Laboratory,
Tampere University of Technology, Finland,
katkov, karen, foi, jta@cs.tut.fi

Introduction

We wish to recover a $2D$ image intensity $y(x)$, from observations

$$z(x) = (y \star v)(x) + \eta(x),$$

where v is the blurring point spread function (PSF) of a linear discrete convolution, $x \in X$,

$$X = \{(k_1, k_2), k_1 = 1, 2, \ldots, n_1, k_2 = 1, 2, \ldots, n_2\}$$

and η is i.i.d. Gaussian noise with the variance σ^2.

The $v(x)$ is known.

The blurring phenomenon, modelled by the kernel v (continuous or discrete), is very evident in many image applications.

Inverse problems arise in optical systems, satellite imaging, radiometry, ultrasonic and magnetic resonance imaging, etc.

In the $2D$ discrete Fourier transform (DFT) domain we have

$$Z(f) = Y(f)V(f) + \eta(f),$$

with $Z(f), Y(f), V(f)$ and $\eta(f)$ of the respective size $n_1 \times n_2$ being the DFT of the corresponding variables and the $2D$ normalized frequency $f \in F$,

$$F = \{(f_1, f_2),\ f_1 = 2\pi k_1/n_1,\ f_2 = 2\pi k_2/n_2,$$
$$k_1 = 0, 1, \ldots, n_1 - 1,\ k_2 = 0, 1, \ldots, n_2 - 1\}.$$

An unbiased estimate of $Y(f)$ can be obtained as a straightforward pure ('naive') inverse solution

$$\hat{Y}(f) = \frac{1}{V(f)} Z(f).$$

In the frequency domain being ill-posed means that $V(f)$ may take zero or close to zero values.

A regularized inverse operator gives instead of the pure inverse

$$\hat{Y}(f) = \frac{V(-f)}{|V(f)|^2 + \varepsilon^2} Z(f),$$

where $\varepsilon > 0$ is a regularization parameter.

This typically produces a reconstruction in which certain features of the original image are ‘smoothed away’ and it seems to be a blurred version of the original.

It would be of interest to obtain sharper reconstructions for objects with edges.

The phenomenon of blurring and the goal of edge recovery have been studied by many researchers over the last few years.

A common point of most methods is that some basis functions are applied for a approximation of the object function $y(x)$ in the form of series with coefficients defined from the observations.

These functions may be Fourier harmonics, eigenfunctions of the convolution operator in *SVD* methods or wavelets in wavelet based decompositions.

LPA estimation

Basically different ideas and methods arise from a nonparametric regression approach. It is assumed that the function $y(x)$ is well approximated by a polynomial in some neighborhood of the point of interest x.

The coefficients of the polynomial fit are found by the weighted least square method. This approximation is used in order to calculate an estimate for the point of interest x called also the ”centre” of the *LPA*.

This pointwise procedure determines a *nonparametric character* of the *LPA* estimation.

Let x be a ”center” (reference point) of the *LPA*. The estimate for the point x_s in the neighborhood of the center x is presented as an expansion:

$$\tilde{y}(x, x_s) = C^T \phi_h(x - x_s), \phi_h(x) = \phi(x/h)$$
$$\phi(x) = (\phi_1(x), \phi_2(x), \ldots, \phi_M(x))^T,$$
$$C = (C_1, C_2, \ldots, C_M)^T,$$

where $\phi(x) \in R^M$ is a vector of linear independent $2D$ polynomials of the powers from 0 up to m, $C \in R^M$ is a vector of parameters of this model.

The conventional quadratic criteria function can be applied in order find C:

$$J_h(x) = \sum_{x_s} w_h(x - x_s)(z(x_s) - \tilde{y}(x, x_s))^2, \; w_h(x) = w(x/h)/h^2,$$

where the window w formalizes the localized fitting in a neighborhood of the centre x.

The scale parameter $h > 0$ determines the ”size” of the neighborhood.

The LPA estimate of $y(x)$ is defined as

$$\hat{y}_h(x) = \tilde{y}(x, x_s)|_{x_s = x} = C^T(x)\phi_h(0),$$

and used for estimate calculation for $x_s = x$ only.

It is a key idea of the pointwise nonparametric estimate design.

Thus, we arrive to a linear discrete kernel estimator defined on the lattice X and given by the kernel $g_h(x)$, $x \in X \subset R^2,$ with the scale (window size) parameter $h > 0$:

$$y_h(x) = (g_h \circledast y)(x)$$

The kernel $g_h(x)$ is defined by the equations

$$g_h(x) = w_h(x)\phi_h^T(0)\Phi_h^{-1}\phi_h(x),$$

$$\Phi_h = \sum_x w_h(x)\phi_h(x)\phi_h^T(x).$$

The following holds for g_h:

(G1) The polynomial smoothness, m vanishing moments,

$$(g_h \circledast x^k)(0) = \delta_{|k|,0}, \ |k| \leq m,$$

where $k = (k_1, k_2)$ is a multi-index,

$$|k| = k_1 + k_2, \ x^k = x_1^{k_1} x_2^{k_2};$$

(G2)

$$\|g_h\|^2 = \sum_x |g_h(x)|^2 \leq Bh^{-b}, \ B, \ b > 0.$$

Then, we say that the g_h is a smoothing kernel estimator of the order m.

Idea of the deblurring algorithm

The smoothed image intensity $y_h(x)$ is used instead of the original $y(x)$ as a solution of the inverse problem and the scale parameter h is exploited in order to suppress noise as much as possible while preserving details of the object function $y(x)$.

Applying the kernel operator g_h to the both sides of the observation equation we yield

$$z_h(x) = g_h(\circledast(y \circledast v))(x) + \eta_h(x) = (y_h \circledast v)(x) + \eta_h(x),$$

where $y_h(x) = (g_h \circledast y)(x)$.

In the frequency domain this equation can be represented as:

$$Z_h(f) = Y_h(f)V(f) + \eta_h(f),$$

where $Z_h(f)$, $Y_h(f)$, and $\eta_h(f)$ stand for DFT of the corresponding smoothed functions.

The following three types of "solutions" can be used:

(A) Pure (naive) inverse (*PI*)

$$\hat{Y}_h(f) = \frac{1}{V(f)} Z_h(f) = \frac{G_h(f)}{V(f)} Z(f), \ V(f) \neq 0, f \in F,$$

(B) Regularized inverse (*RI*)

$$\hat{Y}_h(f) = \frac{V(-f)}{|V(f)|^2 + \varepsilon^2} Z_h(f) = \frac{V(-f) G_h(f)}{|V(f)|^2 + \varepsilon^2} Z(f),$$

(C) Regularized Wiener inverse (*RWI*)

$$\hat{Y}_h(f) = \frac{V(-f)|Y_h(f)|^2}{|V(f)Y_h(f)|^2 + \varepsilon^2 \sigma^2 |G_h(f)|^2} Z_h(f) = \frac{V(-f)|Y(f)|^2 G_h(f)}{|V(f)Y(f)|^2 + \varepsilon^2 \sigma^2} Z(f),$$

where ε is a regularization parameter.

Pointwise spatially adaptive deblurring : ICI rule

The parameter h should be selected in such way that the noise in $\hat{y}_h(x)$ will be suppressed as much as possible provided that the specific features of the object $y(x)$ are preserved in $\hat{y}_h(x)$.

Consider a finite set of scale parameters h:

$$H = \{h_1 < h_2 <\ldots.< h_J\},$$

starting with a quite small $h_1,$ and determine a sequence of the confidence intervals $D(j)$ of the biased estimates obtained with the windows $h = h_j$ as follows

$$D(j) = [\hat{y}_{h_j}(x) - \Gamma \cdot \sigma_{\hat{y}_{h_j}},\ \hat{y}_{h_j}(x) + \Gamma \cdot \sigma_{\hat{y}_{h_j}}],$$

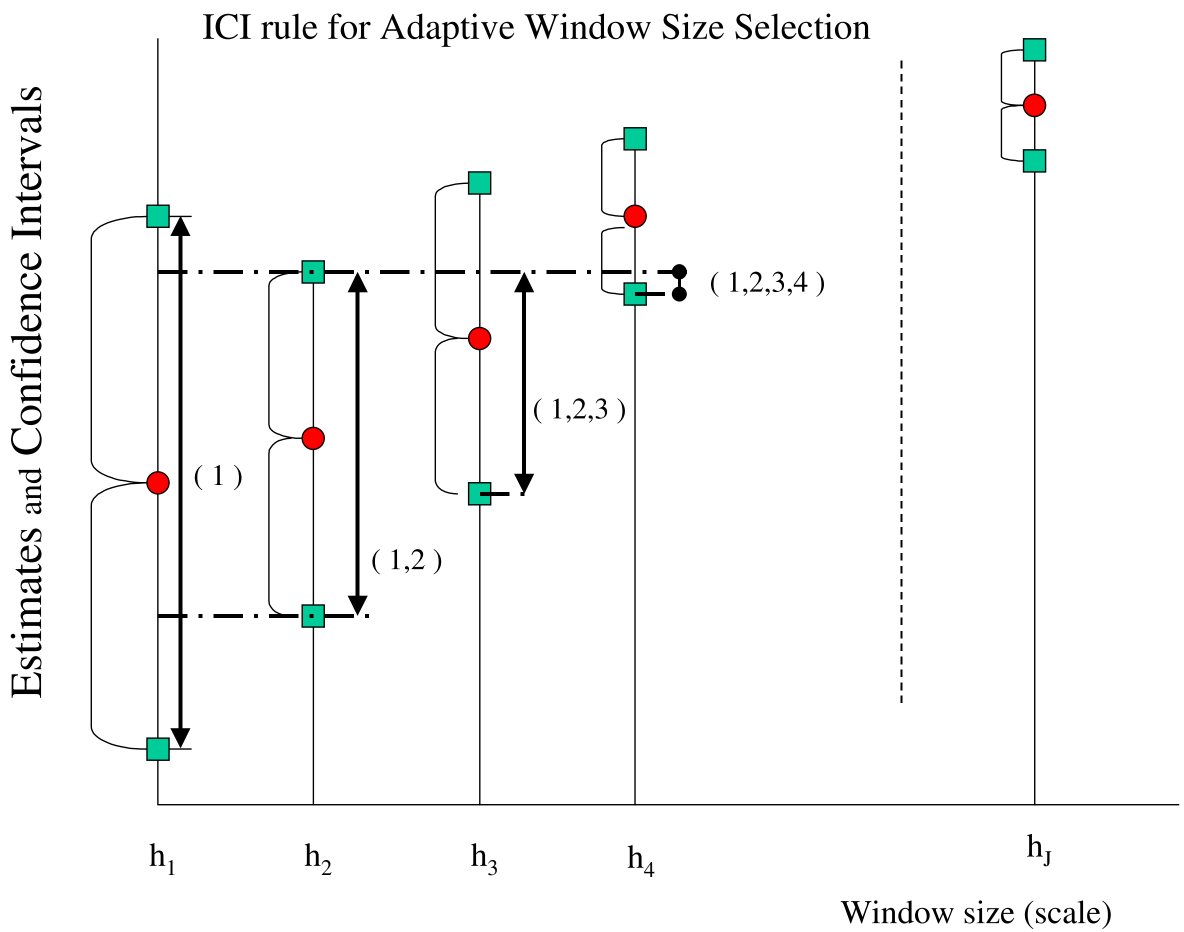
where $\sigma^2_{\hat{y}_h}$ is the variance of the estimate $\hat{y}_{h_j}(x)$ and Γ is a threshold of the confidence interval.

The following *ICI* rule (statistic) is used in order to obtain the adaptive window size (Goldenshluger-Nemirovski (1994,1997)):

Consider the intersection of the intervals $D(j),\ 1 \leq j \leq i,$ with increasing i, and let i^{+} be the largest of those i for which the intervals $D(j),\ 1 \leq j \leq i,$ have a point in common.

This i^{+} defines the adaptive window size and the adaptive LPA estimate as follows

$$\hat{y}^{+}(x) = \hat{y}_{h^{+}(x)}(x),\ h^{+}(x) = h_{i^{+}}.$$



ICI rule for Adaptive Window Size Selection
Estimates and Confidence Intervals
(1)
(1,2)
(1,2,3)
(1,2,3,4)
h1
h2
h3
h4
hJ
Window size (scale)

This window size *ICI* selection procedure requires knowledge of the estimate and its variance only.

It is equally applicable to all three algorithms *PI*, *RI*, *RWI* with the their variances defined respectively by the formulas:

$$\sigma^2_{\hat{y}_h} = \frac{\sigma^2}{n_1 n_2} \| \frac{G_h(f)}{V(f)} \|_2^2,$$

$$\sigma^2_{\hat{y}_h} = \frac{\sigma^2}{n_1 n_2} \| \frac{V(-f) G_h(f)}{|V(f)|^2 + \varepsilon^2} \|_2^2,$$

$$\sigma^2_{\hat{y}_h} = \frac{\sigma^2}{n_1 n_2} \| \frac{V(-f) |Y(f)|^2 G_h(f)}{|V(f) Y(f)|^2 + \varepsilon^2 \sigma^2} \|_2^2.$$

The kernel operator g_h should be agreed with the *PSF* convolution kernel v, i.e. the above variances are assumed to be finite.

Implementation

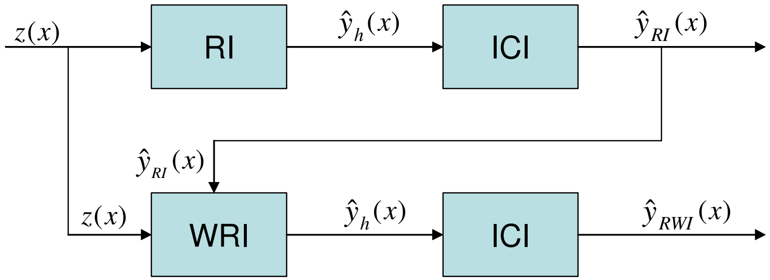
Modification of the basic ideas.

(1) Two step algorithm is developed:

(a) The *RI* deconvolution procedure gives the image estimate used as a reference signal on the second step;

(b) The *RWI* deconvolution gives the final estimate.

Both steps use the *ICI* rule for the adaptive scale selection.



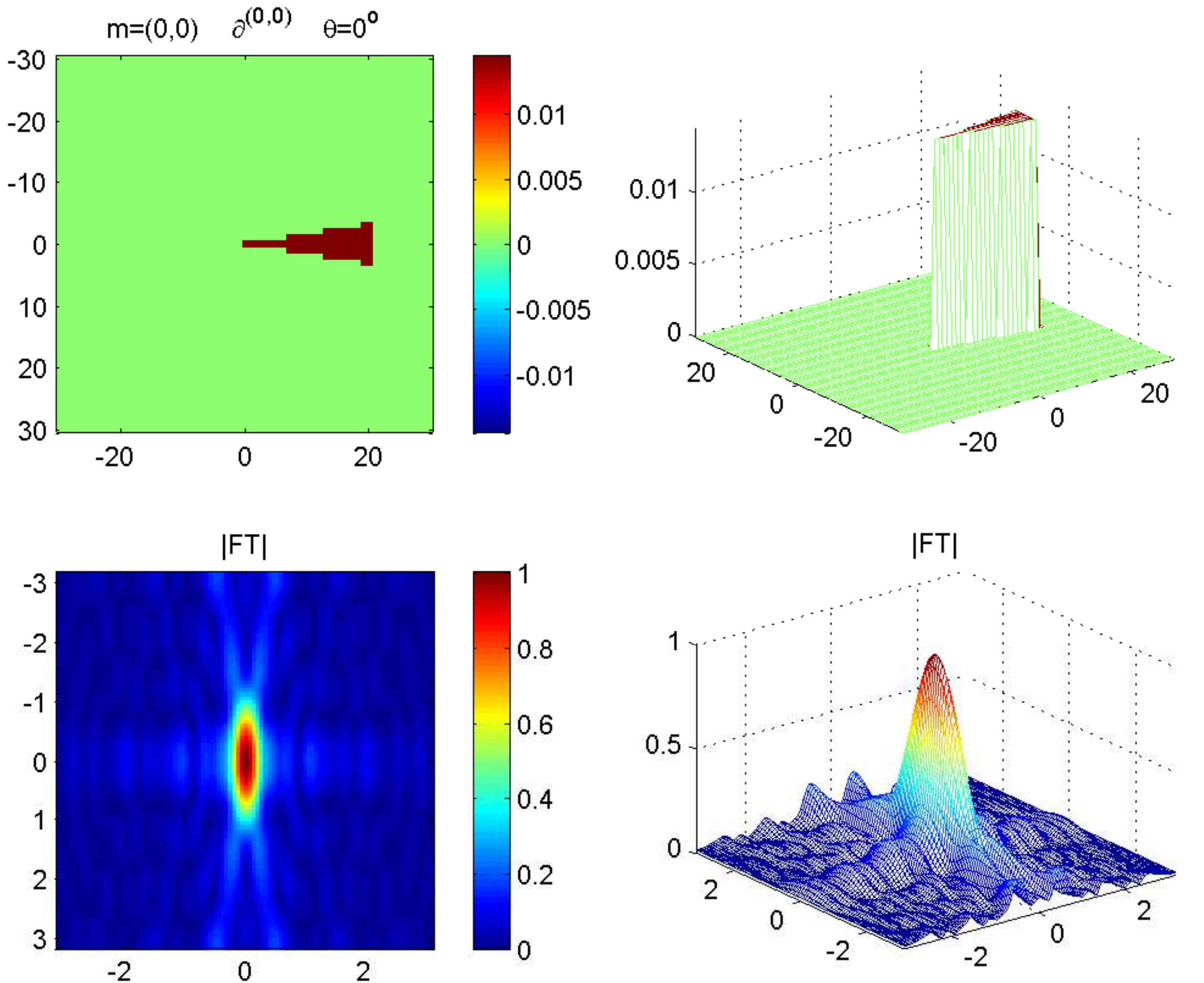
(2) Directional LPA

A symmetric window-weight w in the LPA is a good choice if $y(x)$ is isotropic in a neighborhood of an estimation point. However if $y(x)$ is anisotropic nonsymmetric approximations of $y(x)$ become much more reasonable. To deal with the anisotropy of $y(x)$ multiple *directional LPA* kernels and estimates are exploited.

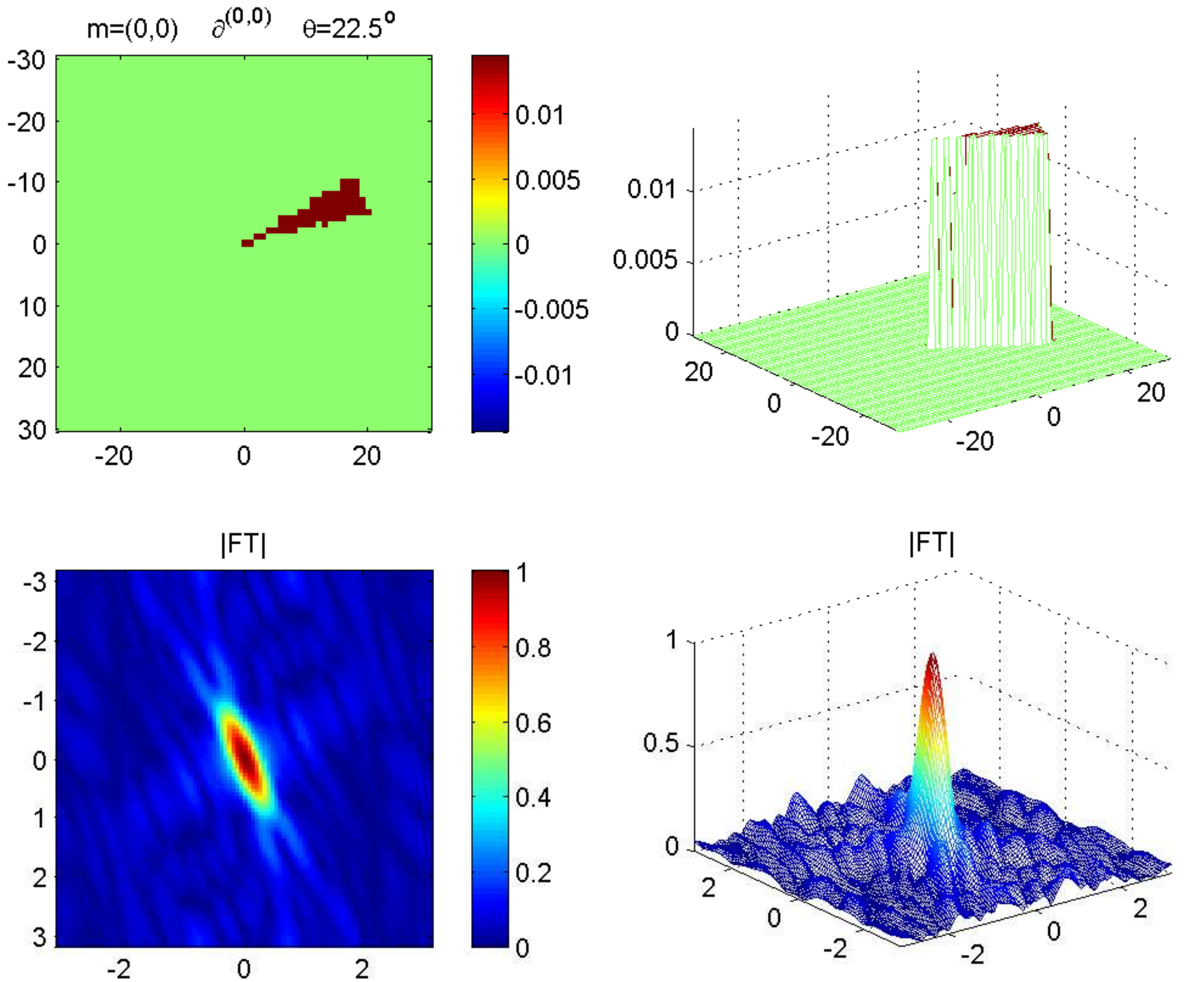
The neighborhood of the pixel x is separated in a number overlapping or nonoverlapping subareas and the narrow directional kernels are obtained by rotation of g_h:

$$g_h(k,\theta) = w_h(Uk)\phi_h^T(0)\Phi_h^{-1}\phi_h(Uk),\ \Phi_h = \sum_k w_h(Uk)\phi_h(Uk)\phi_h^T(Uk),$$

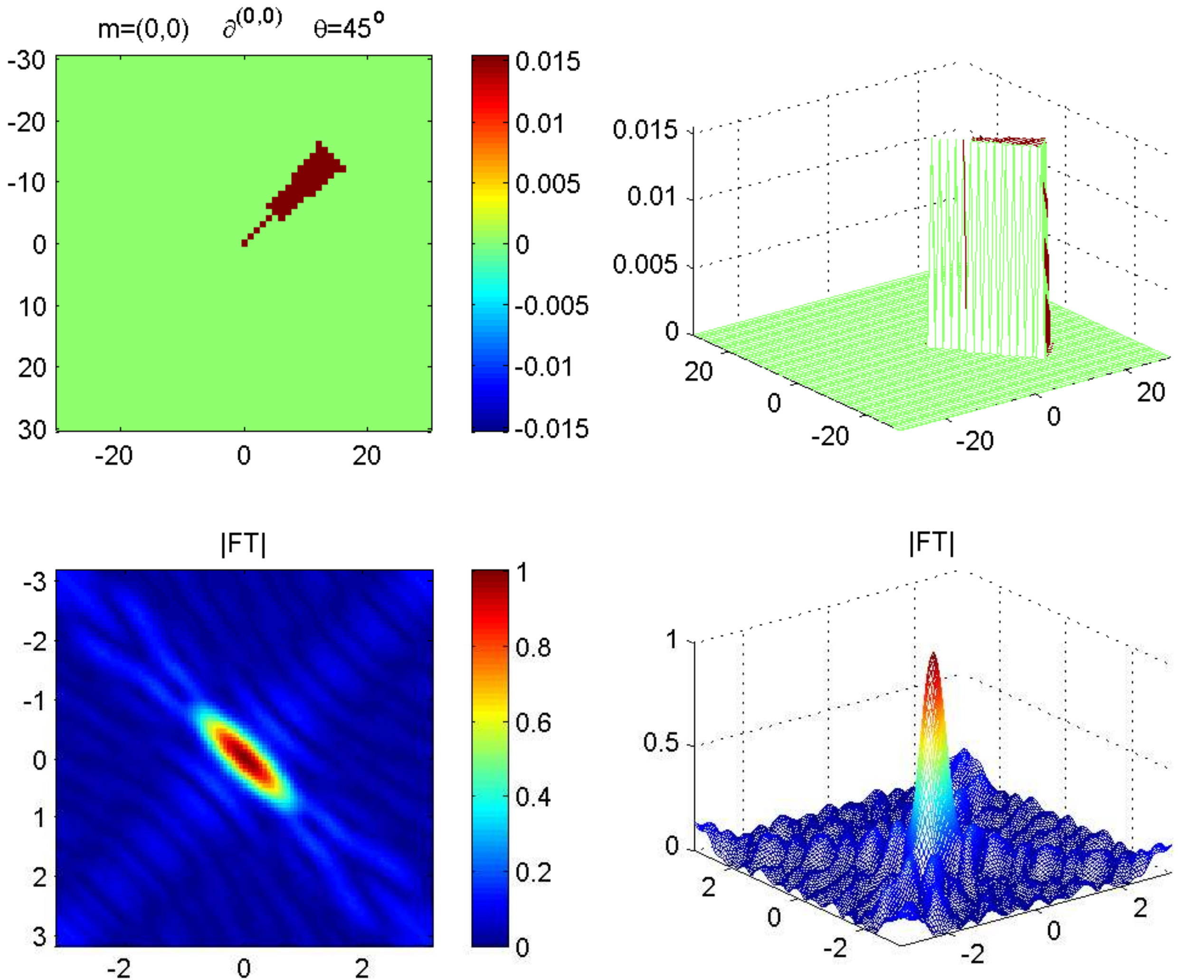
$$U = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

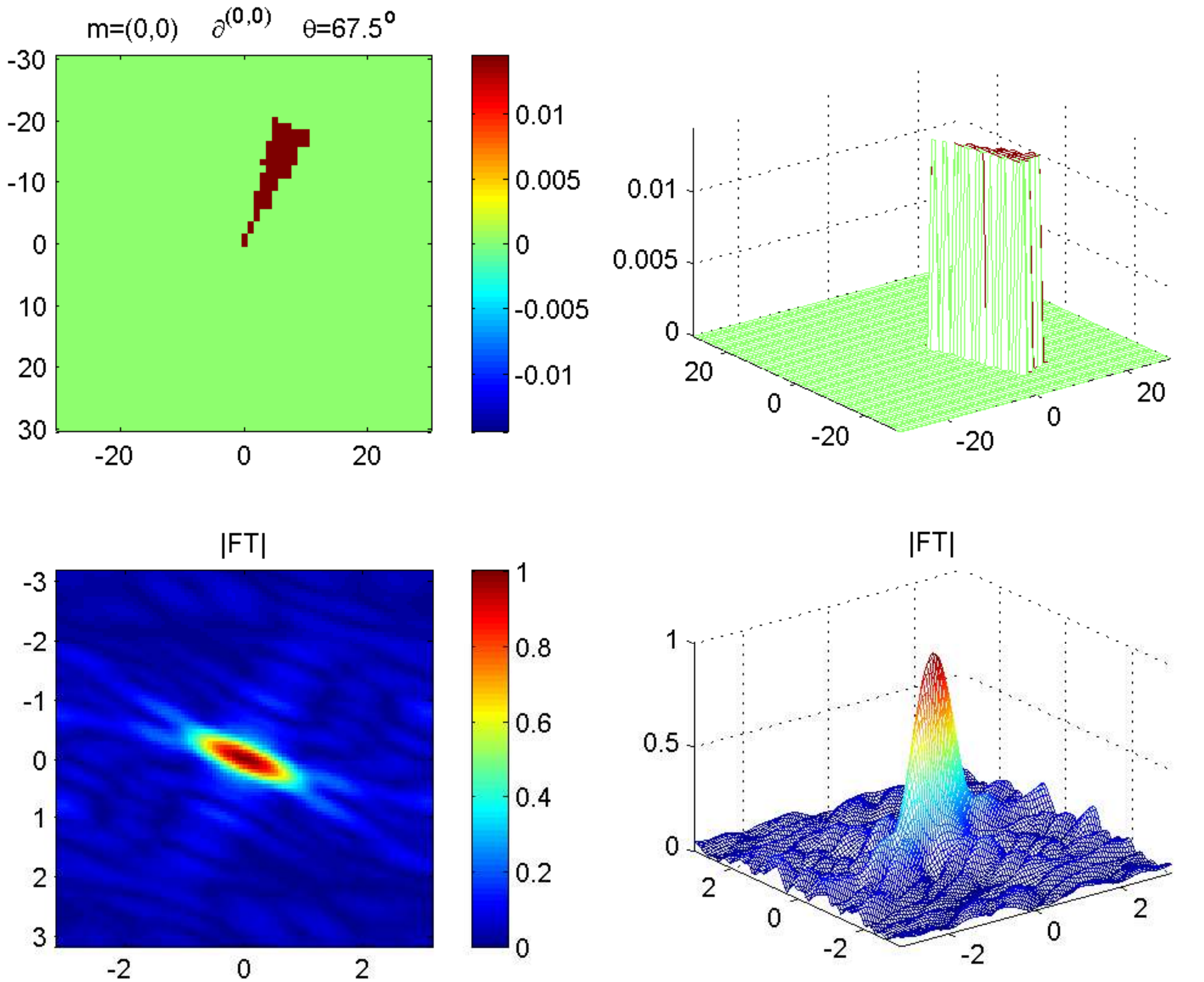


m=(0,0)
$\partial^{(0,0)}$
θ=0°
|FT|
|FT|

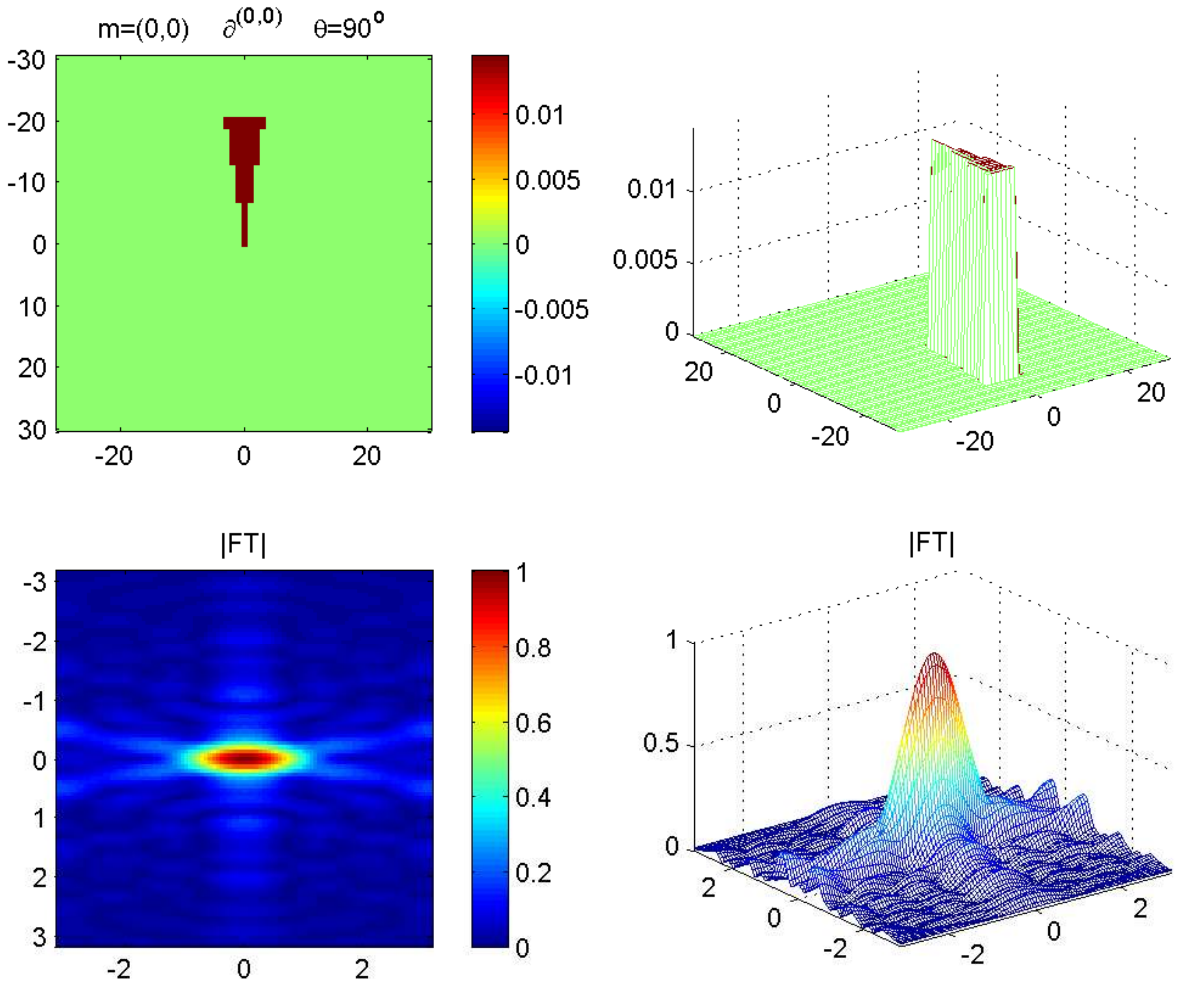


m=(0,0)
$\partial^{(0,0)}$
θ=22.5°
|FT|
|FT|

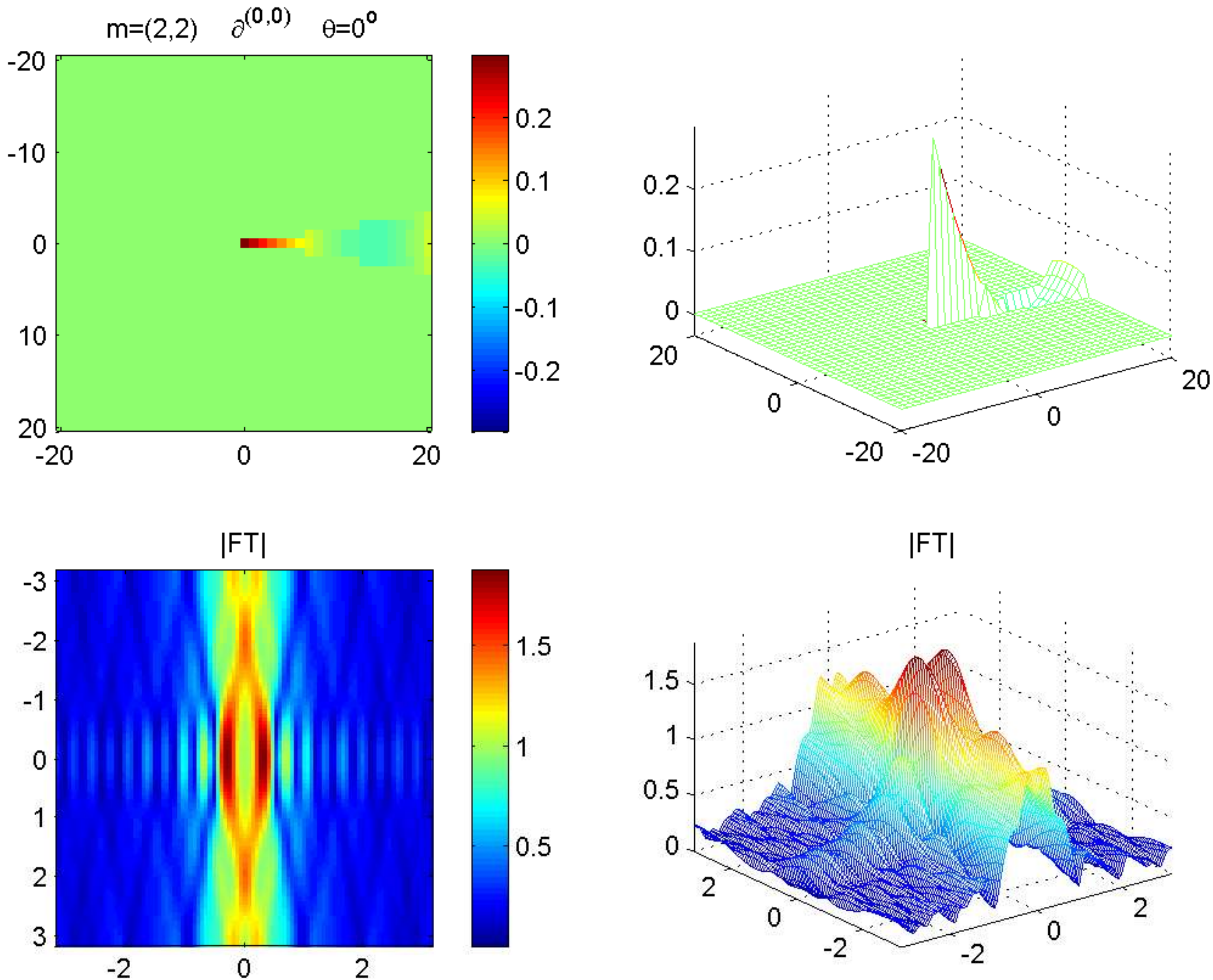

m=(0,0)
∂(0,0)
θ=45°
-30
-20
-10
0
10
20
30
-20
0
20
0.015
0.01
0.005
0
-0.005
-0.01
-0.015
0.015
0.01
0.005
0
20
0
-20
-20
0
20
|FT|
-3
-2
-1
0
1
2
3
-2
0
2
1
0.8
0.6
0.4
0.2
0
|FT|
1
0.5
0
2
0
-2
-2
0
2



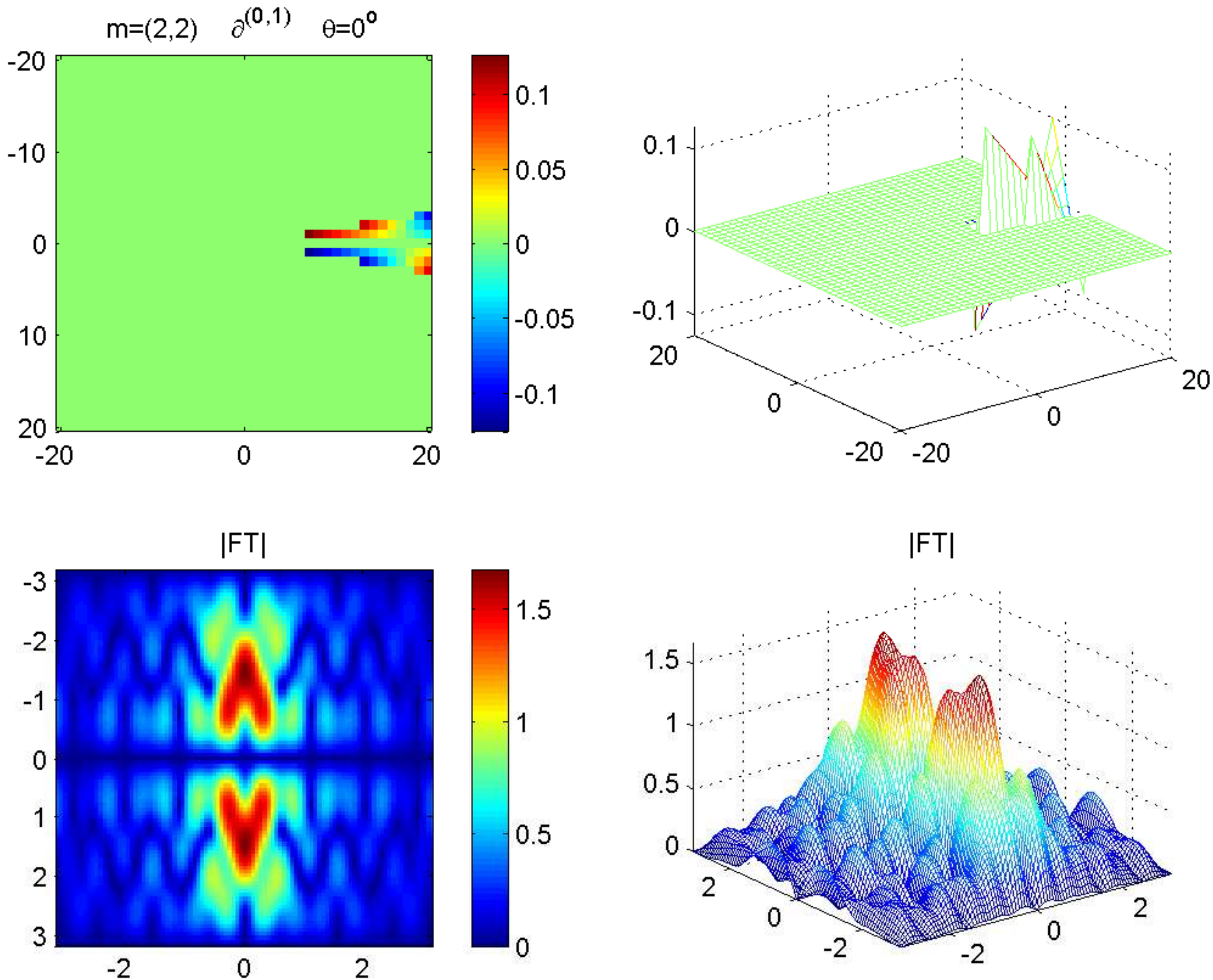
m=(0,0)
$\partial^{(0,0)}$
θ=67.5°
-30
-20
-10
0
10
20
30
-20
0
20
0.01
0.005
0
-0.005
-0.01
0.01
0.005
0
20
0
-20
20
0
-20
|FT|
-3
-2
-1
0
1
2
3
-2
0
2
1
0.8
0.6
0.4
0.2
0
|FT|
1
0.5
0
2
0
-2
2
0
-2



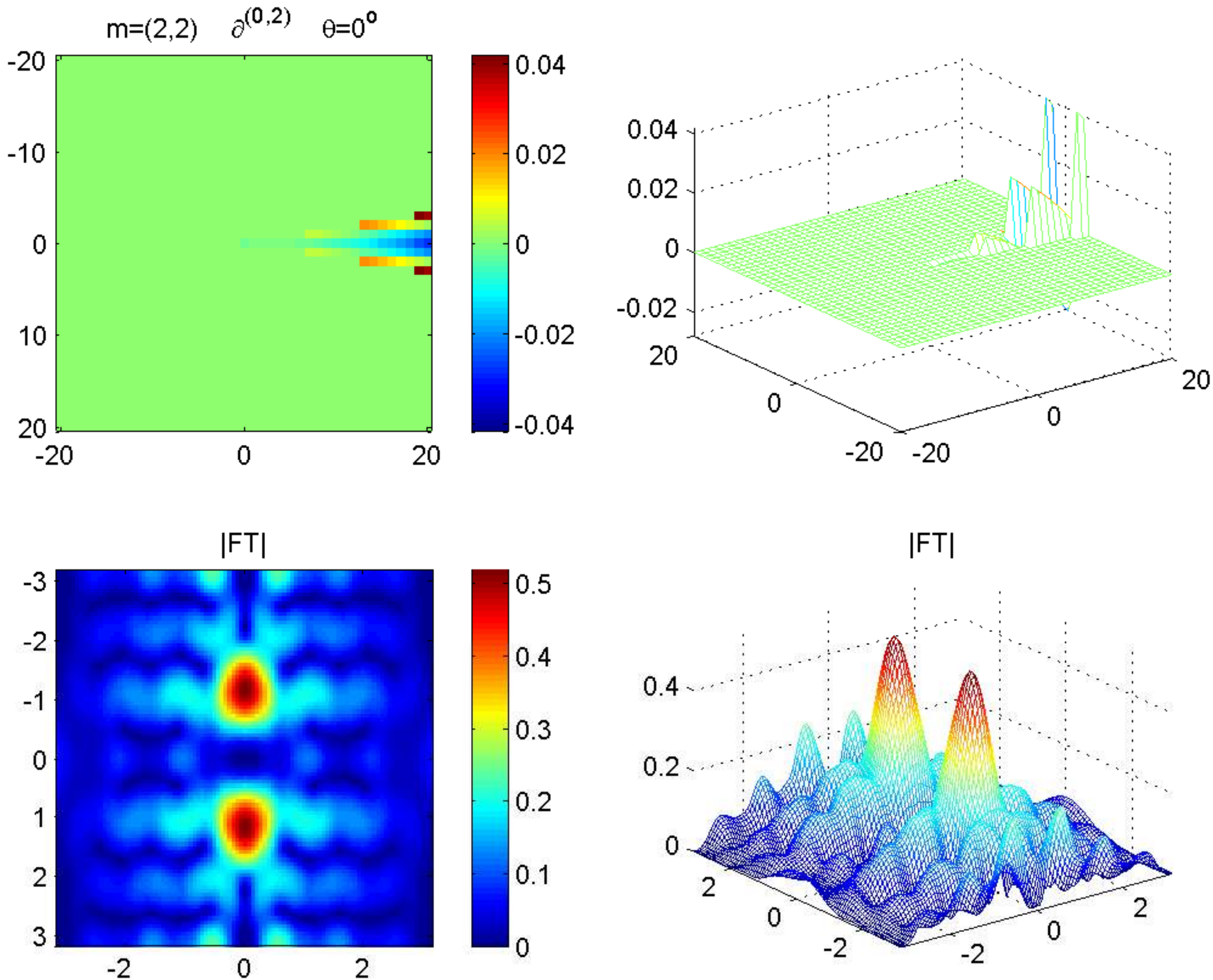
m=(0,0)
$\partial^{(0,0)}$
θ=90°
|FT|
|FT|



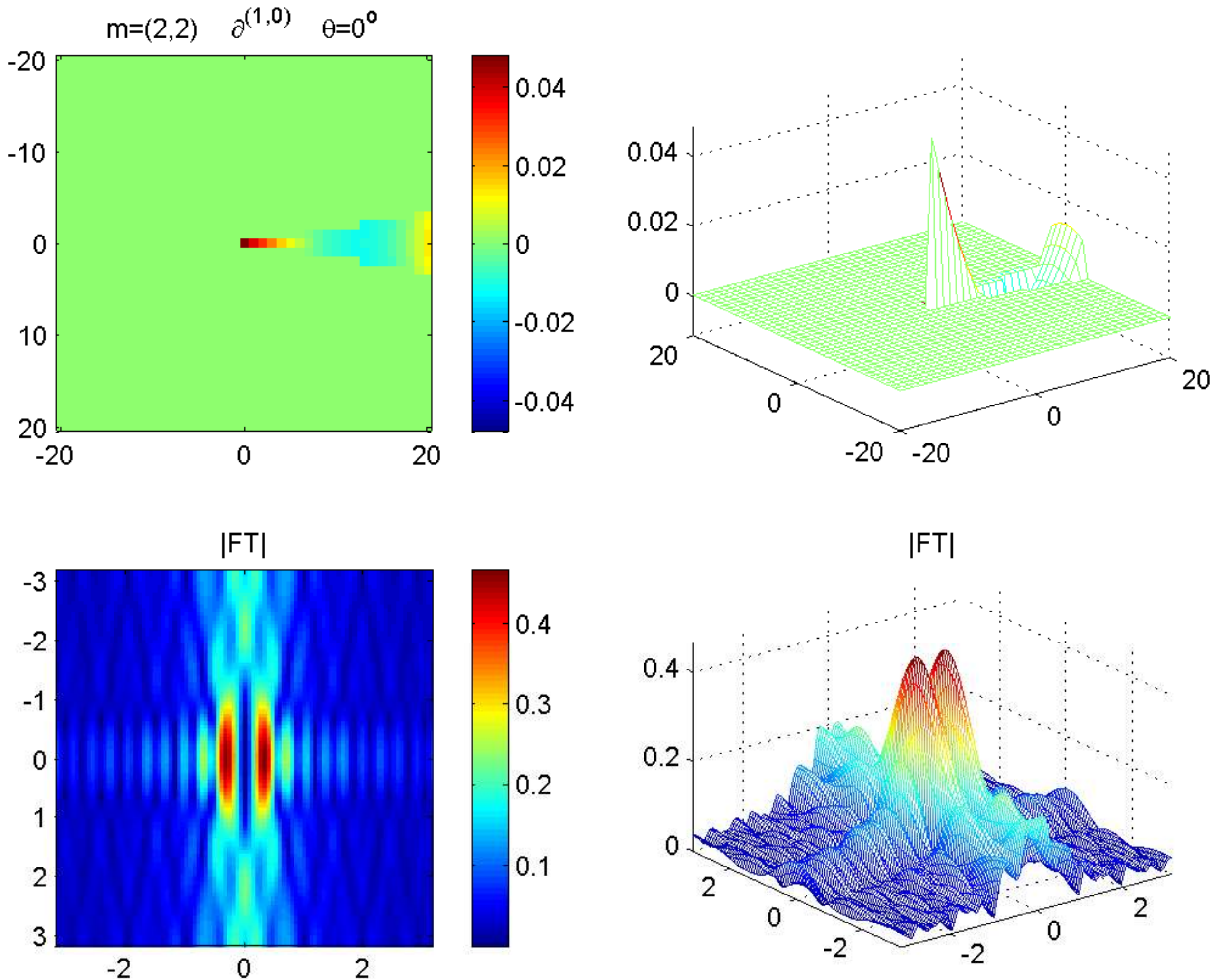
m=(2,2)
$\partial^{(0,0)}$
θ=0°
|FT|
|FT|



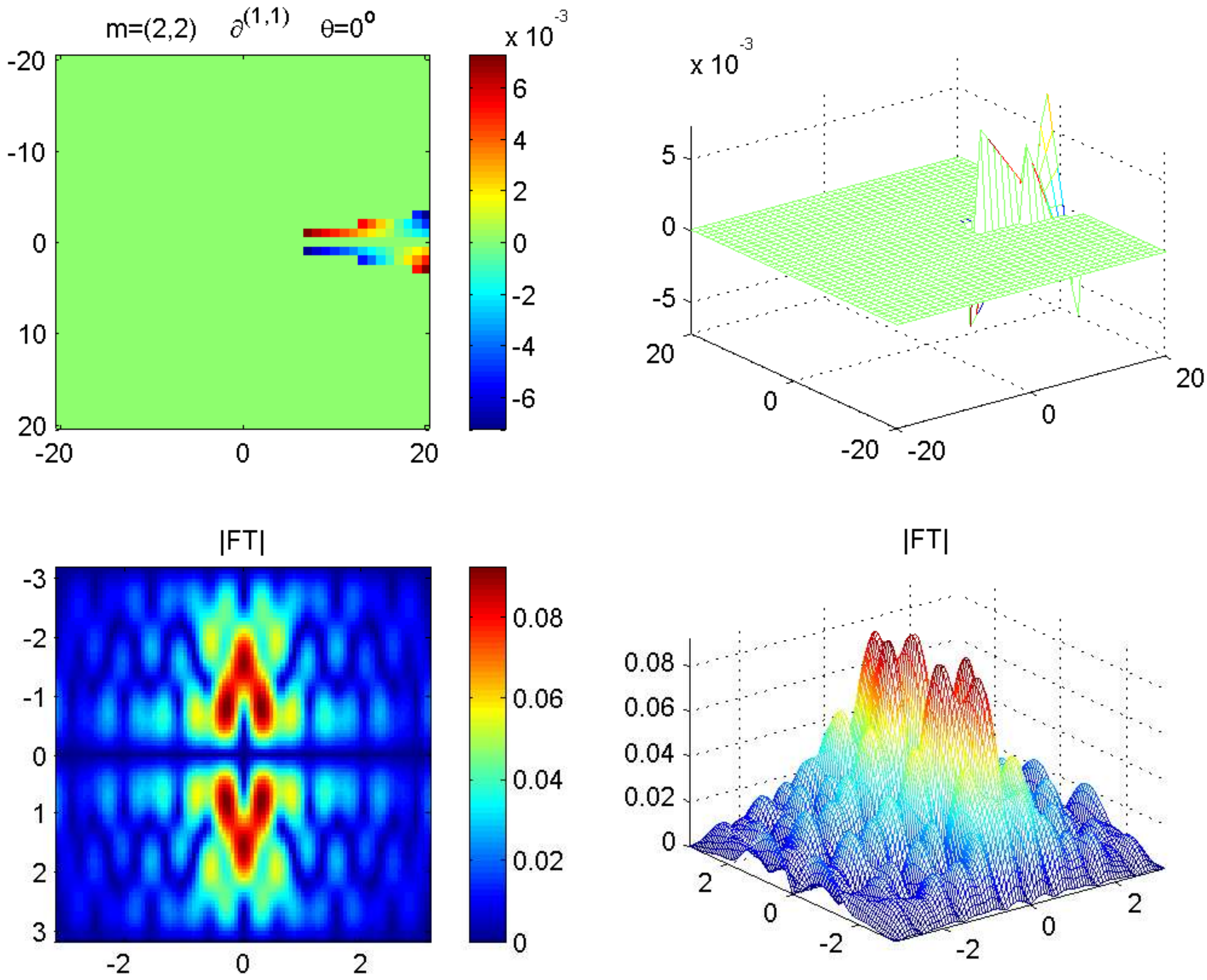
m=(2,2)
$\partial^{(0,1)}$
θ=0°
|FT|
|FT|



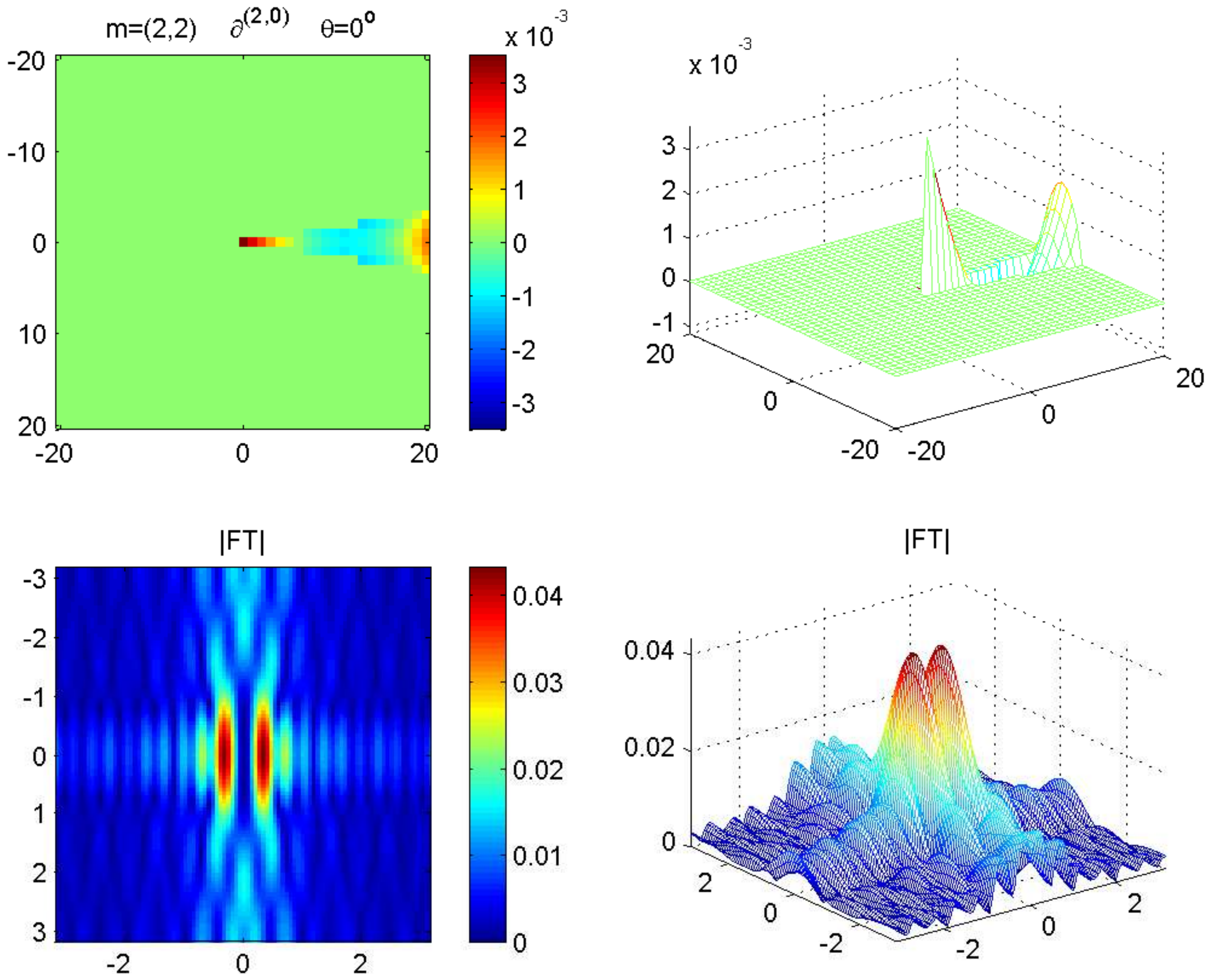
m=(2,2) $\partial^{(0,2)}$ θ=0°
|FT|
|FT|



m=(2,2)
$\partial^{(1,0)}$
θ=0°
|FT|
|FT|



m=(2,2)
$\partial^{(1,1)}$
θ=0°
x 10^-3
|FT|
|FT|


m=(2,2)
$\partial^{(2,0)}$
θ=0°
x 10^-3
|FT|
|FT|

(2) ICI rule

Filtering of $h^{+}(x)$.

(3) *Fusing of partial estimates*

Let's S partial estimates be obtained:

$$\hat{y}_h^{[s]}(x), s = 1, \ldots, S.$$

There are a number of ways how to fuse these estimates. In particular:

$$\hat{y}(x) = \sum_{k=1}^{S} \lambda_s \hat{y}_{h_s^+(x)}^{[s]}(x),\ \lambda_s = \frac{var}{var_s},\ var = 1/\sum_{s=1}^{S}(1/var_s),$$

where $\hat{y}_{h_s^+(x)}^{[s]}(x)$ are the kernel estimates with the ICI rule adaptive window size for the subareas $s = 1, \ldots, S$.

Algorithm complexity

The calculation of the image estimate $\hat{y}_h^{[s]}$ for given h is a linear convolution requiring $N_{conv} \sim n \log n$ operations $n = n_1 n_2$.

If the sectioning procedure is used for convolution, then $N_{conv} \sim n \log n_{h_J}$, where n_{h_J} is a maximum size of the square mask of the kernel g_{h_J}.

The selection of the adaptive scale the ICI algorithm is implemented as a loop on J different scales in the set H.

These calculations are repeated for each of the S subareas (quadrants) of the pixel neighborhood with the fusing the estimates.

Thus, overall the algorithm complexity is proportional to

$$J \bullet S \bullet N_{conv},$$

where S is a number of the directional estimates.

Experiments

The test signals are the 256×256 "Cameraman" image (8 bit gray-scale) and binary "Cheese" corrupted by an additive zero-mean Gaussian noise.

The blurred SNR ($BSNR$) is defined in dB as

$$BSNR = 10 \log_{10} [\sum_x \{(y \circledast v)(x) - mean((y \circledast v)(x))\}^2 / \sigma^2 n_1 n_2],$$

with $BSNR = 40\ dB$.

The discrete-space blur convolution PSF is a uniform 9×9 box-car. The LPA is defined with $H = \{1, 3, 5, 9, 17\}$.

The performance criteria:

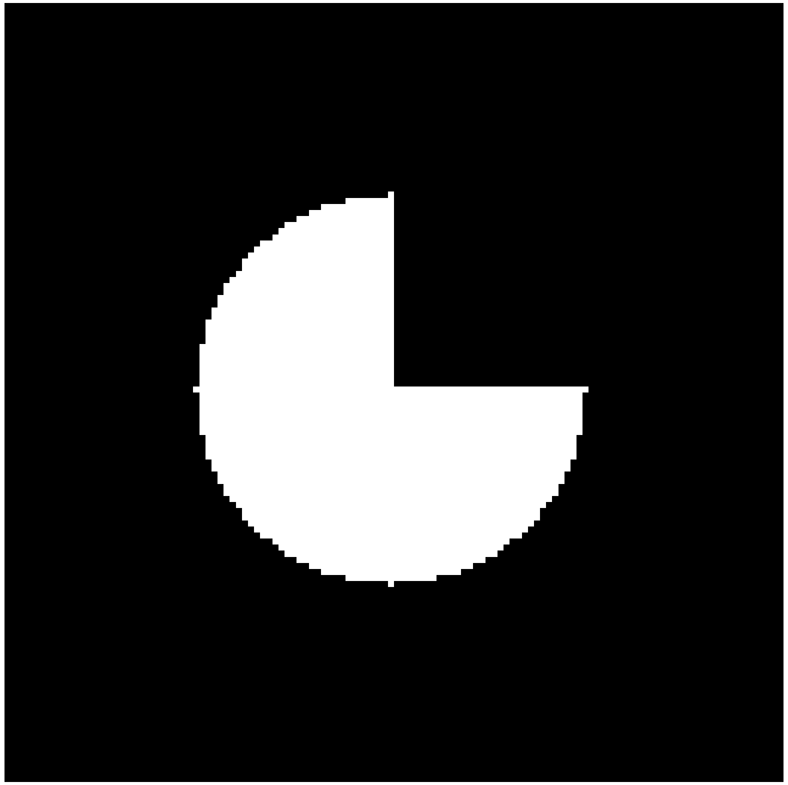
(1) Root mean squared error (*RMSE*): $RMSE = \sqrt{\frac{1}{\#}\sum_x (y(x) - \hat{y}(x))^2}$;

(2) Improvement in *SNR* (*ISNR*) in *dB*:

$$ISNR = 20\log_{10}\left(\sqrt{\frac{1}{\#}\sum_x (y(x) - z(x))^2} / RMSE\right).$$

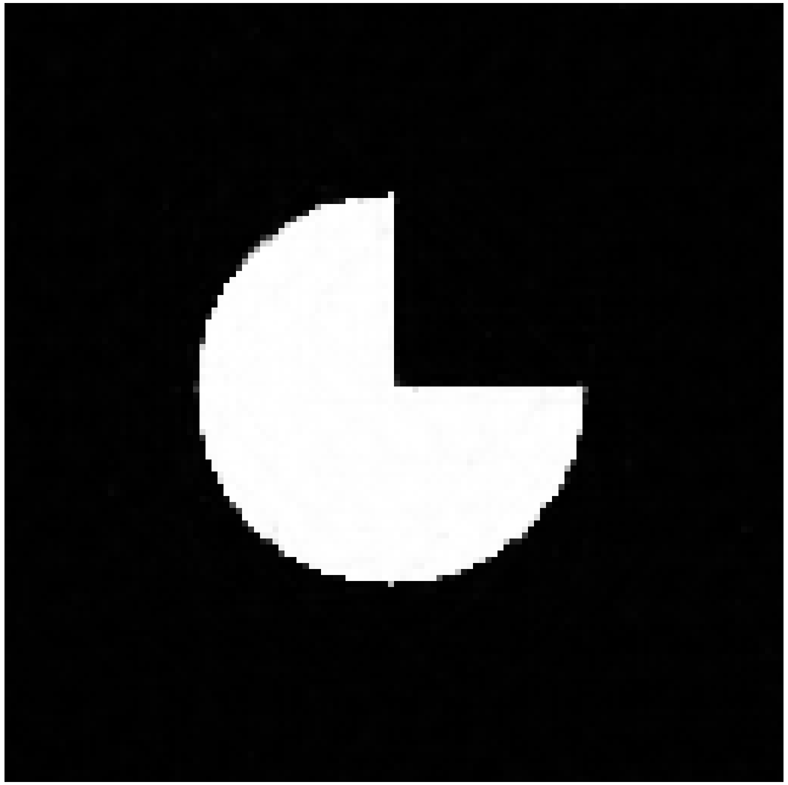
(3) Visual evaluation.

Mainly the comparison is produced with one of the best in the field wavelet based algorithm developed in RICE University (Neelamani R., Choi H., Baraniuk R.). It is called ”*ForWaRD*” and available from *http://www.dsp.rice.edu/software/*.

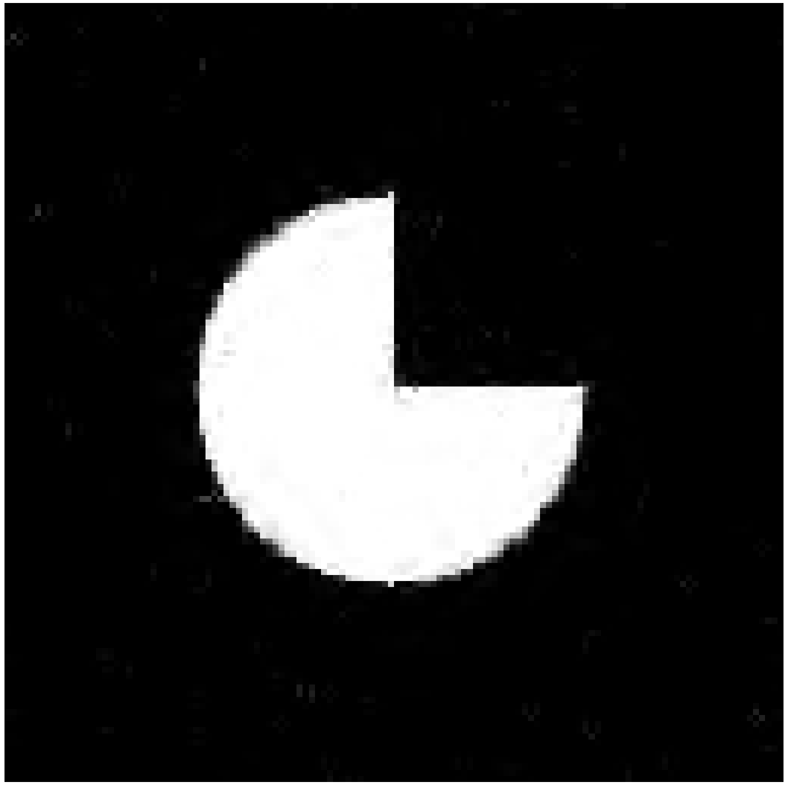


"Cheese" original

"Cheese" observations

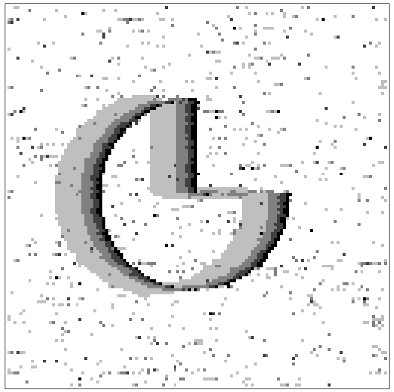


$LPA - ICI$, $ISNR = 15.8dB$

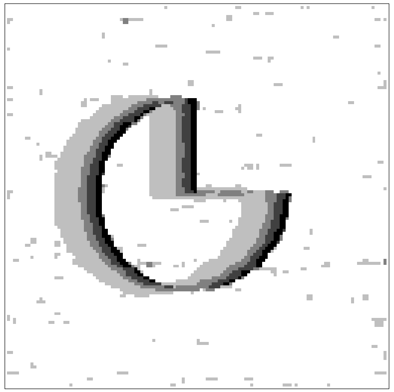


$ForWard,\ ISNR = 9.55dB$

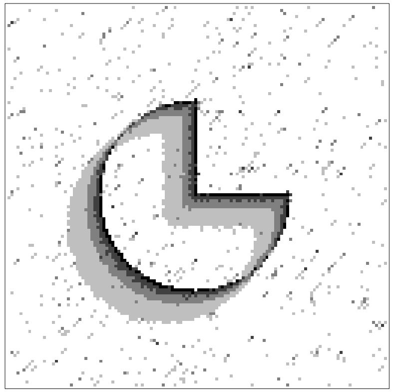
$h_+(x)$, no filt., $\theta = 0$



$h_+(x)$, filt., $\theta = 0$



$h_+(x)$, no filt., $\theta = 45^0$



$h_+(x)$, filt., $\theta = 45^0$

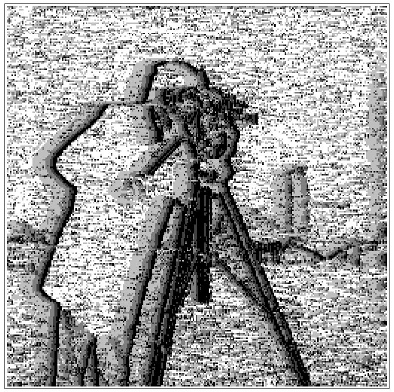
Original

Observations

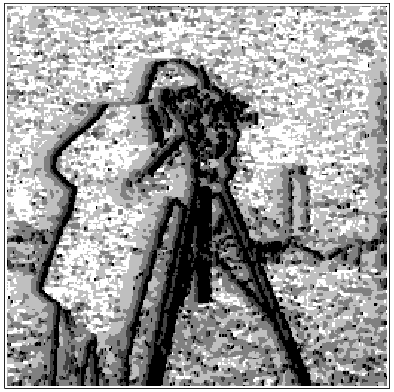
$LPA - ICI$, $ISNR = 8.23dB$

$ForWard$, $ISNR = 7.16dB$

h_+, no filt. $\theta = 0$



h_+, filt. $\theta = 0$



h_+, no filt. $\theta = 45^0$

h_+, filt. $\theta = 45^0$

TABLE I : ”Cameraman”, 9×9 car-box PSF

Method	ISNR
$LPA-ICI$	8.23 dB
Result by Figueiredo and Nowak (EM)	7.59 dB
$Forward$,	7.16 dB
Result by Banham and Katsaggelos	6.7 dB

TABLE II : "Lena”, PSF is a 5×5 separable filter with the weights $[1, 4, 6, 4, 1]/16$ in horizontal and vertical directions, $BSNR = 15.93$ dB.

Method	ISNR
$LPA - ICI$	3.76 dB
Best result by Figueiredo and Nowak (EM)	2.94 dB
Best result by Liu and Moulin	1.078 dB
$ForWaRD$	2.87 dB

Conclusions

Some theoretical topics:

(1) Directional nonparametric regression estimation;

(2) The adaptive scale selection procedures for small number of samples;

(3) Fusing of the partial directional estimates;

(4) Analysis of regularized and Wiener based algorithms.

Asymptotic properties

Oracle accuracy:

(B1) $y \in H_r$

$$H_r = \{\max_{r_1+r_2=r} |\partial^{r_1+r_2} y(x)/\partial x_1^{r_1} \partial x_2^{r_2}| = L_r(x) \leq \bar{L}_r, \ \forall \ r_1 + r_2 = r\}.$$

(B2) $V(\lambda)$ is polynomial decaying,

$$\bar{c}_0 \geq |V(\lambda)| \|\lambda\|^\alpha \geq c_0, \ \forall \ \|\lambda\| > A, \min_{\|\lambda\| \leq A} |V(\lambda)| = c_1.$$

Then, as $h, \Delta, \Delta/h \to 0, \ m \leq r$,

$$r(x, h^*(x)) \ = O(n^{-\frac{2(m+1)}{m+\alpha+2}}).$$

Adaptive estimate accuracy:

$$r(x, h^+(x)) \ = O((\ln n/n)^{\frac{2(m+1)}{m+\alpha+2}}).$$

can be proved based on the technique due to A.Goldenshluger, A.

Nemirovski (1997) for $1D$ regression and A. Goldenshluger (1999) for $1D$ continuous deconvolution.

References

Banham M. R. and A.K. Katsaggelos, ”Spatially adaptive wavelet-based multiscale image restoration,” IEEE Trans. Image Processing, vol. 5, pp. 619-634, 1996.

A. Goldenshluger, ”On pointwise adaptive nonparametric deconvolution,” Bernoulli, vol. 5, pp. 907-925, 1999.

A. Goldenshluger and A. Nemirovski, ”On spatial adaptive estimation of nonparametric regression,” Math. Meth. Statistics vol. 6, pp. 135-170, 1997.

V. Katkovnik, Nonparametric identification and smoothing of data (Local approximation methods). Nauka, Moscow, 1985 (in Russian).

V. Katkovnik, “A new method for varying adaptive bandwidth selection,” IEEE Trans. on Signal Processing, vol. 47, N^{0}9, pp. 2567-2571, 1999.

V. Katkovnik, K. Egiazarian, and Jaakko Astola, “Adaptive window size image de-noising based on intersection of confidence intervals (ICI) rule,” Journal of Mathematical Imaging and Vision, vol. 16, N 3. pp. 223-235, 2002.

V. Katkovnik, K. Egiazarian, and Jaakko Astola, Adaptive varying scale methods in image processing. Tampere International Center for Signal Processing, TICSP Series, N 19, Tampere, TTY, Monistamo, 2003.

V. Katkovnik, K. Egiazarian, and Jaakko Astola, "A spatially adaptive nonparametric regression image deblurring," IEEE Trans Image Processing, submitted, 2003.

Katkovnik, V., ”A multiresolution nonparametric regression and image de-nosing,” Proceedings of ICIP 2003, Barcelona, Spain, 14-17 September, 2003.

V. Katkovnik, K. Egiazarian, and Jaakko Astola, "Spatially adaptive deblurring based on nonparametric regression ," ICIP 2004, Singapore, submitted, 2004.

V. Katkovnik, A. Foi, K. Egiazarian, and J. Astola, "Directional varying scale approximations for anisotropic signal processing", EUSIPCO 2004, Wien, submitted, 2004.

R. Neelamani, H. Choi, and R. G. Baraniuk, ”Forward: Fourier-wavelet regularized deconvolution for ill-conditioned

systems,” IEEE Trans. on Image Processing, 2003.

Figueiredo M. A.T. and R. D. Nowak,” An EM algorithm for wavelet-based image restoration,” IEEE Trans Image Processing, vol. 12, no. 8, pp. 906-916, 2003.