UNIVERSITÄT BERN

Total Variation Blind Deconvolution: The Devil is in the Details*

Paolo Favaro
Computer Vision Group
University of Bern

*Joint work with Daniele Perrone

Engineering Science — University of Oxford

Blur in pictures

- When we take a picture we expose the sensor of our camera to the incoming light through the lens

- The lens needs to be placed at the right distance between the scene and the sensor, otherwise…

Out of focus blur

Blur in pictures

- When we take a picture we expose the sensor of our camera to the incoming light through the lens

- The camera or the scene should not move during the exposure otherwise…

Motion Blur

A blur model

- When the captured image is blurry then we have no choice but to try and remove the degradation computationally

- The first step is to model blur degradation

$$f = k * u + n$$

blurry image | kernel | sharp image | noise

Deblurring

- When the kernel k is known then we are essentially inverting a linear system

- Deblurring can be posed as a convex optimization problem

$$\min_u \lambda \|u\|_{BV} + \frac{1}{2}\|f - k * u\|_2^2$$

Kernel k is known: Deblurring

Kernel k is known: Deblurring

Blind deconvolution

- Neither the kernel nor the sharp image are known
- We need to recover both the blur and the sharp image

$$\min_{u,k} \lambda \|u\|_{BV} + \frac{1}{2}\|f - k * u\|_2^2$$

- The problem is non convex

Blind deconvolution

- Neither the kernel nor the sharp image are known
- We need to recover both the blur and the sharp image

$$\min_{u,k} \lambda \|u\|_{BV} + \frac{1}{2}\|f - k * u\|_2^2$$

- The problem is non convex

Prior Work

- Before 1996-1998 the general belief was that blind deconvolution was not just impossible, but that it was hopelessly impossible

- How can we extract more data than we observe?

Ambiguities

- The main difficulty in solving blind deconvolution is that the problem is ill-posed

Ambiguities

- The main difficulty in solving blind deconvolution is that the problem is ill-posed
- For example, if (u,k) is a solution, then also (au,k/a) and (u(x+d),k(x-d)) for any d and for any a>0 are solutions

Ambiguities

- The main difficulty in solving blind deconvolution is that the problem is ill-posed
- For example, if (u,k) is a solution, then also (au,k/a) and (u(x+d),k(x-d)) for any d and for any a>0 are solutions
- Consider the Fourier transform: F = KU, where F, K and U are Fourier transforms of f, k, and u respectively

Ambiguities

- The main difficulty in solving blind deconvolution is that the problem is ill-posed
- For example, if (u,k) is a solution, then also (au,k/a) and (u(x+d),k(x-d)) for any d and for any a>0 are solutions
- Consider the Fourier transform: F = KU, where F, K and U are Fourier transforms of f, k, and u respectively
- Then, for any K that is not 0 at any frequency there exists always a U such that F = KU (simply let U = F/K)

The role of the image prior

- To reduce the set of ambiguities to a unique sensible answer one can use a regularization term

- One of the first regularization terms proposed in blind deconvolution was the H^1 prior (You and Kaveh 1996)

$$\|\nabla u\|_2^2$$

- Total Variation (strongly related to sparse gradient and natural image prior) was also proposed at the same time (You and Kaveh 1996)

$$\|\nabla u\|_2$$

Chan and Wong (1998)

- Total Variation Blind Deconvolution (similar work appeared earlier in You and Kaveh, 1996)
- Solve

$$\min_{u,k} \lambda \|u\|_{BV} + \frac{1}{2}\|f - k * u\|_2^2$$

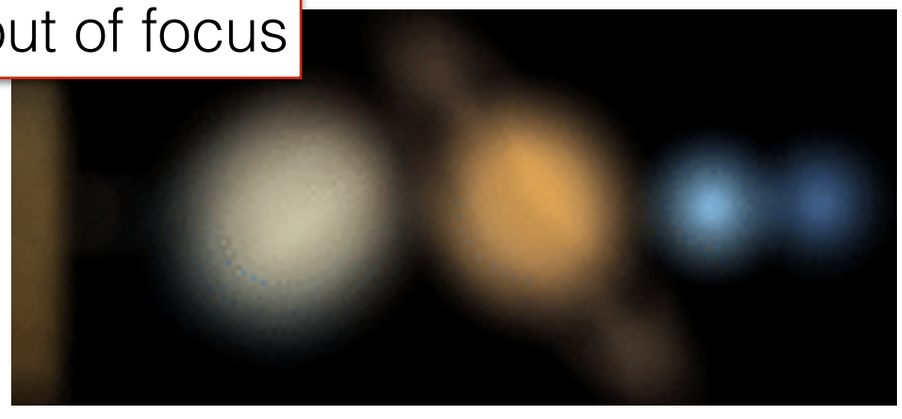
- Use an alternating minimization algorithm (fix the blur and compute the sharp image, then fix the sharp image and compute the blur)

Chan and Wong (1998)

it works!

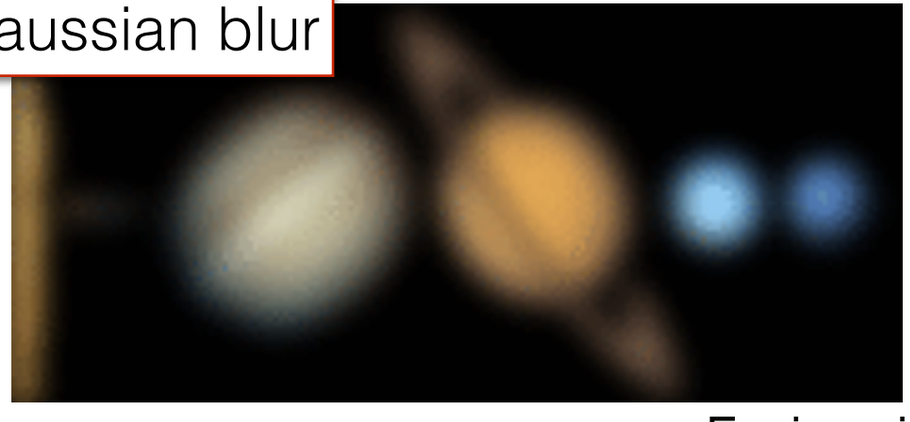
sharp image

out of focus

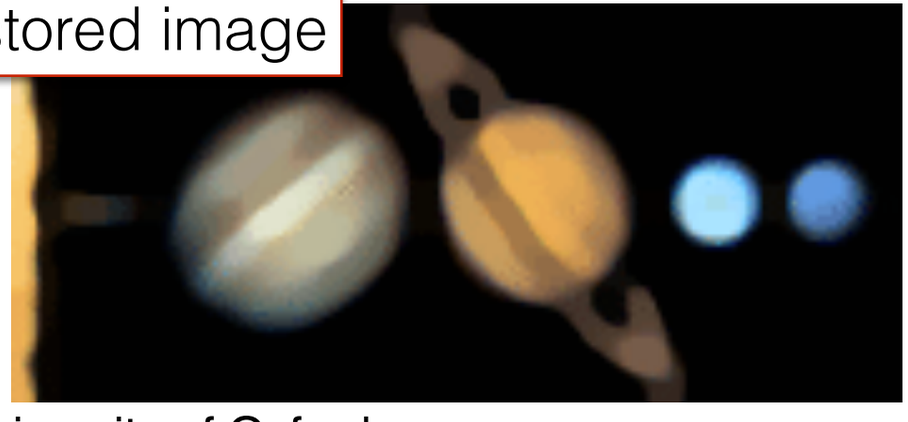


restored image

Gaussian blur



restored image



Fergus et al (2006)

- Alternating minimization does not work ($MAP_{u,k}$)

A straightforward approach to deconvolution is to solve for the maximum a-posteriori (MAP) solution, which finds the kernel $\mathbf{K}$ and latent image gradients $\nabla\mathbf{L}$ that maximizes $p(\mathbf{K}, \nabla\mathbf{L}_p|\nabla\mathbf{P})$. This is equivalent to solving a regularized-least squares problem that attempts to fit the data while also minimizing small gradients. We tried this (using conjugate gradient search) but found that the algorithm failed. One interpretation is that the MAP objective function attempts to minimize all gradients (even large ones), whereas we expect natural images to have some large gradients. Consequently, the algorithm yields a two-tone image, since virtually all the gradients are zero. If we reduce the noise variance (thus increasing the weight on the data-fitting term), then the algorithm yields a delta-function for $\mathbf{K}$, which exactly fits the blurred image, but without any deblurring. Additionally, we find the MAP objective function to be very susceptible to poor local minima.

Fergus et al (2006)

- Alternating minimization does not work ($MAP_{u,k}$)

A straightforward approach to deconvolution is to solve for the maximum a-posteriori (MAP) solution, which finds the kernel $\mathbf{K}$ and latent image gradients $\nabla\mathbf{L}$ that maximizes $p(\mathbf{K}, \nabla\mathbf{L}_p|\nabla\mathbf{P})$. This is equivalent to solving a regularized-least squares problem that attempts to fit the data while also minimizing small gradients. We tried this (using conjugate gradient search) but found that the algorithm failed. One interpretation is that the MAP objective function attempts to minimize all gradients (even large ones), whereas we expect natural images to have some large gradients. Consequently, the algorithm yields a two-tone image, since virtually all the gradients are zero. If we reduce the noise variance (thus increasing the weight on the data-fitting term), then the algorithm yields a delta-function for $\mathbf{K}$, which exactly fits the blurred image, but without any deblurring. Additionally, we find the MAP objective function to be very susceptible to poor local minima.

Fergus et al (2006)

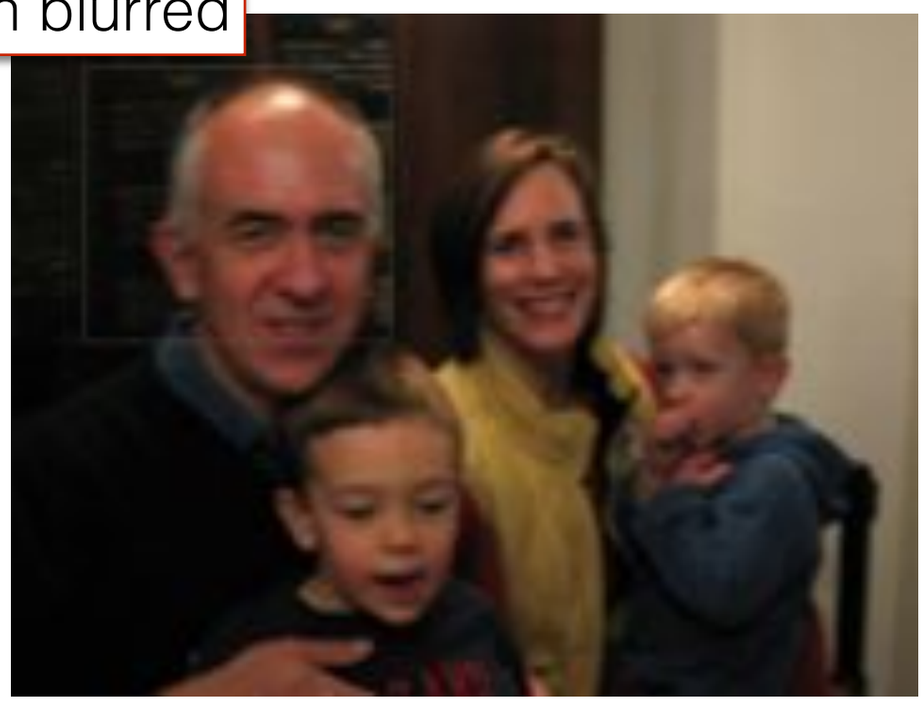
- Alternating minimization does not work ($MAP_{u,k}$)
- Use instead a MAP_k approach (based on Miskin and McKay 2000)
  - Marginalize wrt a distribution of the sharp images
  - Compute k by maximizing the marginalized dist.
  - Compute u by solving deblurring given k
- Technical details: Use a variational bayesian approach (Jordan et al 1999) and a Gaussian mixture model

Fergus et al (2006)

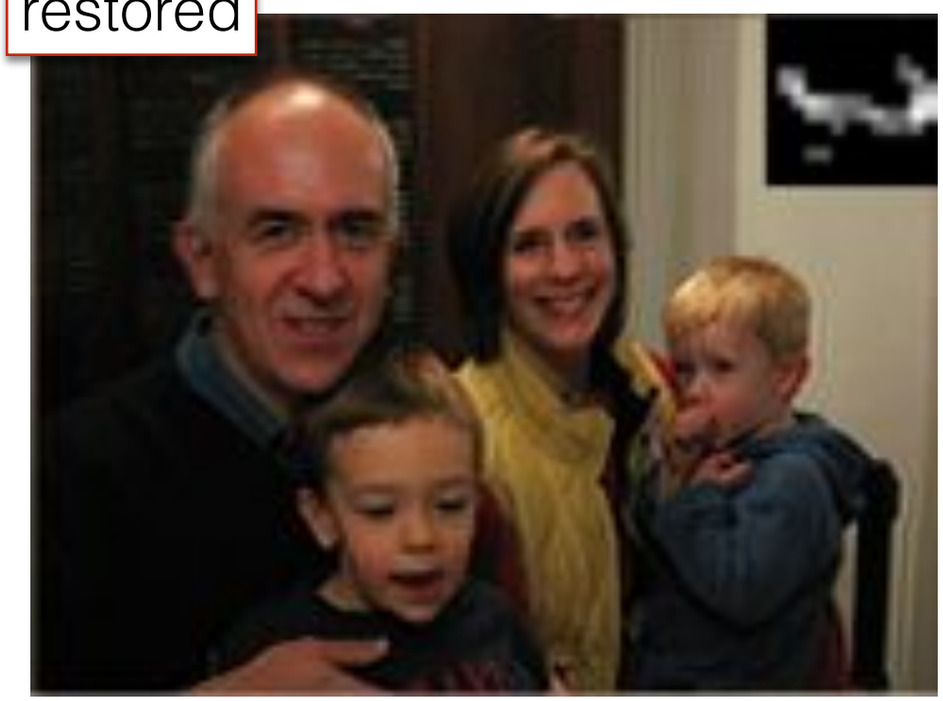
motion blurred

restored

motion blurred



restored



Shan et al (2008)

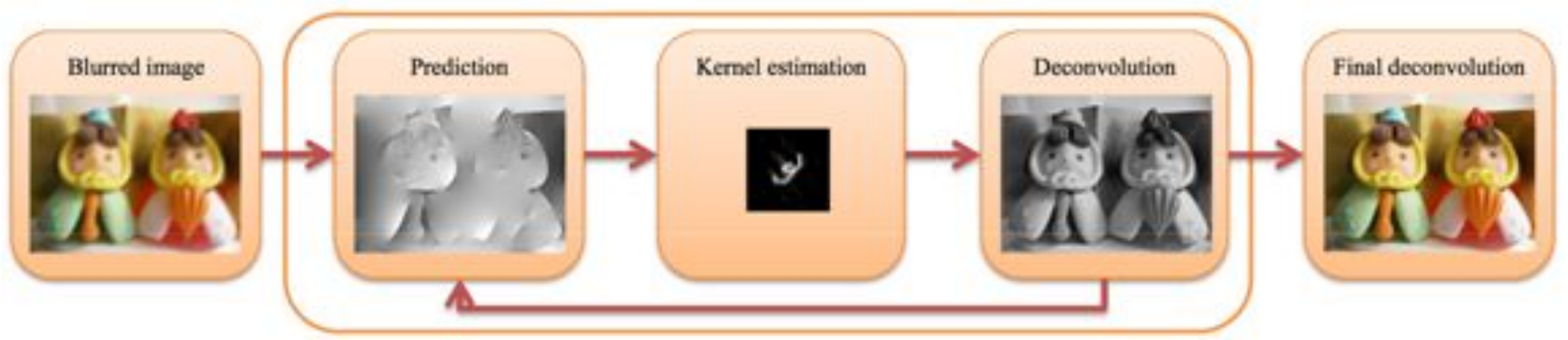
- Impose that noise is iid
- Use alternating minimization ($MAP_{u,k}$) but on the image gradients
- Impose that sharp image and blurry image should coincide where the blurry image is very smooth
- Then estimate sharp image given kernel k

Shan et al (2008)

motion blurred | iteration 1 | iteration 6 | iteration 10

Cho and Lee (2009)

- Success of prior work is: Sharp edge restoration and noise suppression in smooth regions
- Blur can be estimated reliably at edges
- Try and predict edges with a shock filter
- Use a modified alternating minimization ($MAP_{u,k}$)



Cho and Lee (2009)

motion blurred

restored

Xu et al (2013)

- Use a saturated L_1 prior (they call it unnatural L_0)

- Use alternating minimization ($MAP_{u,k}$)

- Technical details: Many intermediate steps

Xu et al (2013)

(a) input (b) Fergus *et al.* [5] (c) Shan *et al.* [20] (d) Cho and Lee [3]

(e) Xu and Jia [25] (f) Hirsch *et al.* [9] (g) Krishnan *et al.* [16] (h) ours

Levin et al (2011)

- Stop using $MAP_{u,k}$! It should not work! Use MAP_k
- Compare the following true solution (u,k) with the no-blur solution (f, δ)

$$f \equiv \delta * f \equiv k * u$$

- Then, solution is based only on the image prior; however, the prior favors the no-blur solution!

$$\|\nabla f\|_2 \leq \|\nabla u\|_2$$

Levin et al (2011)

Claim 2. *Let $g_x(x^0), g_y(x^0)$ be infinitely large gradient images*[2] *sampled from the prior*

$$p^0(g_{x,i}(x^0), g_{y,i}(x^0)) \propto e^{-|g_{x,i}(x^0)|^\alpha - |g_{y,i}(x^0)|^\alpha}. \quad (7)$$

Let $y = k^* \otimes x^0$ with k^* a unit sum kernel whose L_2 norm is small $\|k^*\|^2 \ll 1$. Then, solving $\mathrm{MAP}_{x,k}$ subject to a unit sum constraint on k will fail to recover the true blur kernel. In particular, if k^* is a unit sum kernel with ℓ uniform nonzero values, $\mathrm{MAP}_{x,k}$ subject to a unit sum constraint is guaranteed to fail for large ℓ values.

MAP_k

- After marginalization Levin et al. 2011 obtain

$$\nabla x_{n+1} = \arg\min_x \frac{1}{\eta_n^2}\|\nabla y - x \star k_n\|^2 + \sum_i (w_{i,n} x)^2$$

$$k_{n+1} = \arg\min_k \|\nabla y - \nabla x_{n+1} \star k\|^2 + \lambda_n \|k\|^2 \ ,$$

- which is an alternating minimization
- weights are updated sequentially

A conundrum

- On the one hand many $MAP_{u,k}$ implementations and (heuristic) variants work very well, and on the other hand they are <u>not</u> supposed to work at all

A conundrum

- On the one hand many $MAP_{u,k}$ implementations and (heuristic) variants work very well, and on the other hand they are <u>not</u> supposed to work at all

- Rather than developing yet another blind deconvolution algorithm, should we not try to understand what is going on first?

A conundrum

- On the one hand many $MAP_{u,k}$ implementations and (heuristic) variants work very well, and on the other hand they are <u>not</u> supposed to work at all

- Rather than developing yet another blind deconvolution algorithm, should we not try to understand what is going on first?

- Could MAP_k be just another recipe for $MAP_{u,k}$?

Recent analysis

- Wipf and Zhang arxiv 2013: MAP_k equivalent to a $\text{MAP}_{u,k}$

MAP_k

$$\max_{\mathbf{k}} p(\mathbf{k}|\mathbf{y}) \equiv \min_{\mathbf{k}} -2\log p(\mathbf{y}|\mathbf{k})p(\mathbf{k})$$

$\text{MAP}_{u,k}$

$$\mathcal{L}(\mathbf{x},\mathbf{k}) \triangleq \frac{1}{\lambda}\|\mathbf{y}-\mathbf{k}*\mathbf{x}\|_2^2 + \sum_i g_{\text{VB}}(x_i,\rho) + m\log\|\bar{\mathbf{k}}\|_2^2,$$

$$g_{\text{VB}}(x_i,\rho) \triangleq \min_{\gamma_i \geq 0}\left[\frac{x_i^2}{\gamma_i} + \log(\rho+\gamma_i) + f(\gamma_i)\right], \text{ s.t. } \rho = \frac{\lambda}{\|\bar{\mathbf{k}}\|_2^2}$$

- See also Babacan et al. 2012 and Krishnan et al. 2014

Recent analysis

- So, current conclusion is that it is not about MAP_k vs $MAP_{u,k}$, but rather about the choice of priors

- Still, this does not explain why current so-called $MAP_{u,k}$ approaches (that use TV-like priors) work

Removing the bells and whistles

- We start by applying the golden rule in analysis: Remove all the unnecessary

- Result: Total Variation Blind Deconvolution (1996!)

$$\begin{aligned} \min_{u,k} \quad & \lambda J(u) + \|k * u - f\|_2^2 \\ \text{subject to} \quad & k \succcurlyeq 0, \quad \|k\|_1 = 1 \end{aligned}$$

Attempt #1: Exact solution

- The alternating minimization (AM) algorithm

$$u^{t+1} \leftarrow \min ||k^t * u - f||_2^2 + \lambda J(u)$$

$$\begin{aligned} k^{t+1} \leftarrow \quad & \min_k && ||k * u^t - f||_2^2 \\ & \text{subject to} && k \succcurlyeq 0, \quad \|k\|_1 = 1 \end{aligned}$$

- Actually, it does not work!

AM does not work

Theorem 3.1 *Let* $J(u) = \|\nabla u\|_p \doteq \left(\int \|\nabla u(\mathbf{x})\|_p^p d\mathbf{x}\right)^{\frac{1}{p}}$, *with* $p \in [1, \infty]$, f *be the noise-free input blurry image* ($n = 0$) *and* u^0 *the sharp image. Then,*

$$J(f) \leq J(u^0). \tag{7}$$

Proof. Because f is noise-free, $f = k^0 * u^0$; since the convolution and the gradient are linear operators, we have

$$J(f) = \|\nabla(k^0 * u^0)\|_p = \|k^0 * \nabla u^0\|_p \tag{8}$$

By applying Young's inequality [1] (see Theorem 3.9.4, pages 205-206) we have

$$J(f) = \|k^0 * \nabla u^0\|_p \leq \|k^0\|_1 \|\nabla u^0\|_p = \|\nabla u^0\|_p \doteq J(u^0) \tag{9}$$

since $\|k^0\|_1 = 1$. ■

A toy example in 1D

- Let us consider a 1D signal (a hat function) and a 1D blur of 3 pixels

A toy example in 1D

- Let us consider a 1D signal (a hat function) and a 1D blur of 3 pixels

- Because the blur components add to 1, we only have 2 free parameters

A toy example in 1D

- Let us consider a 1D signal (a hat function) and a 1D blur of 3 pixels

- Because the blur components add to 1, we only have 2 free parameters

- For each possible combination of these parameters we minimize the TV problem wrt the sharp image (a deblurring problem)

A toy example in 1D

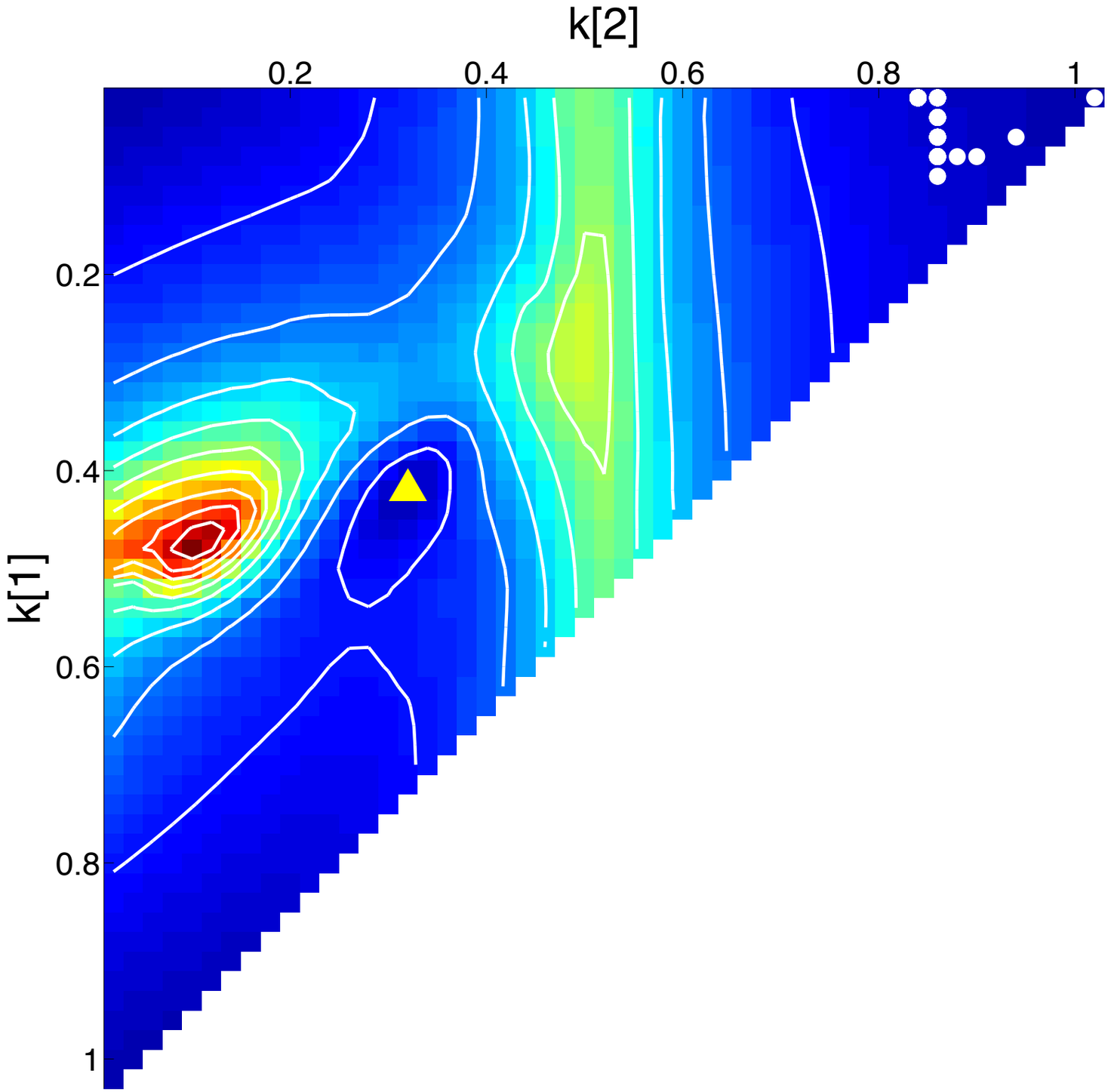
- Let us consider a 1D signal (a hat function) and a 1D blur of 3 pixels

- Because the blur components add to 1, we only have 2 free parameters

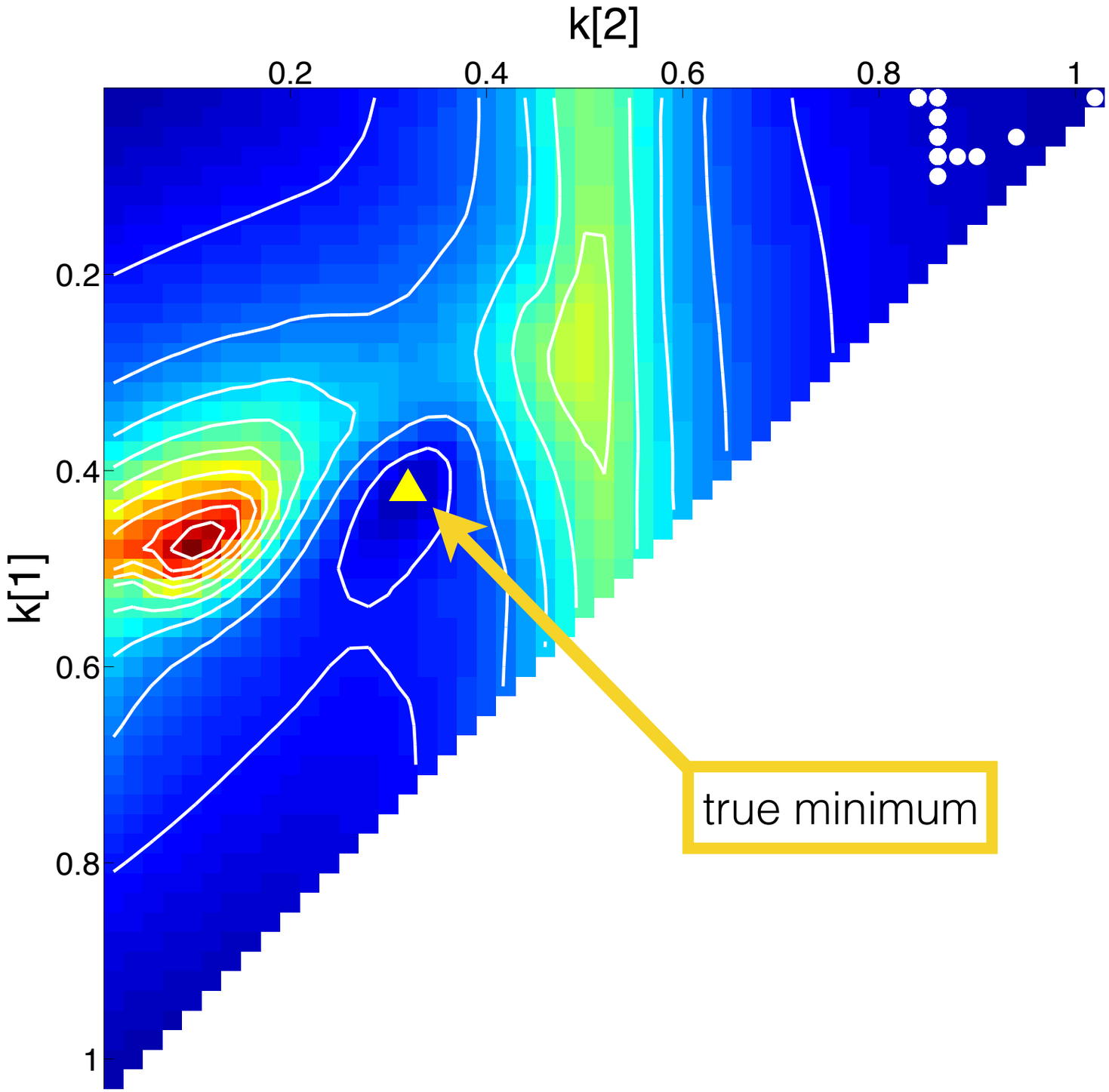
- For each possible combination of these parameters we minimize the TV problem wrt the sharp image (a deblurring problem)

- We show the energy at the minimum wrt the sharp image for each possible blur

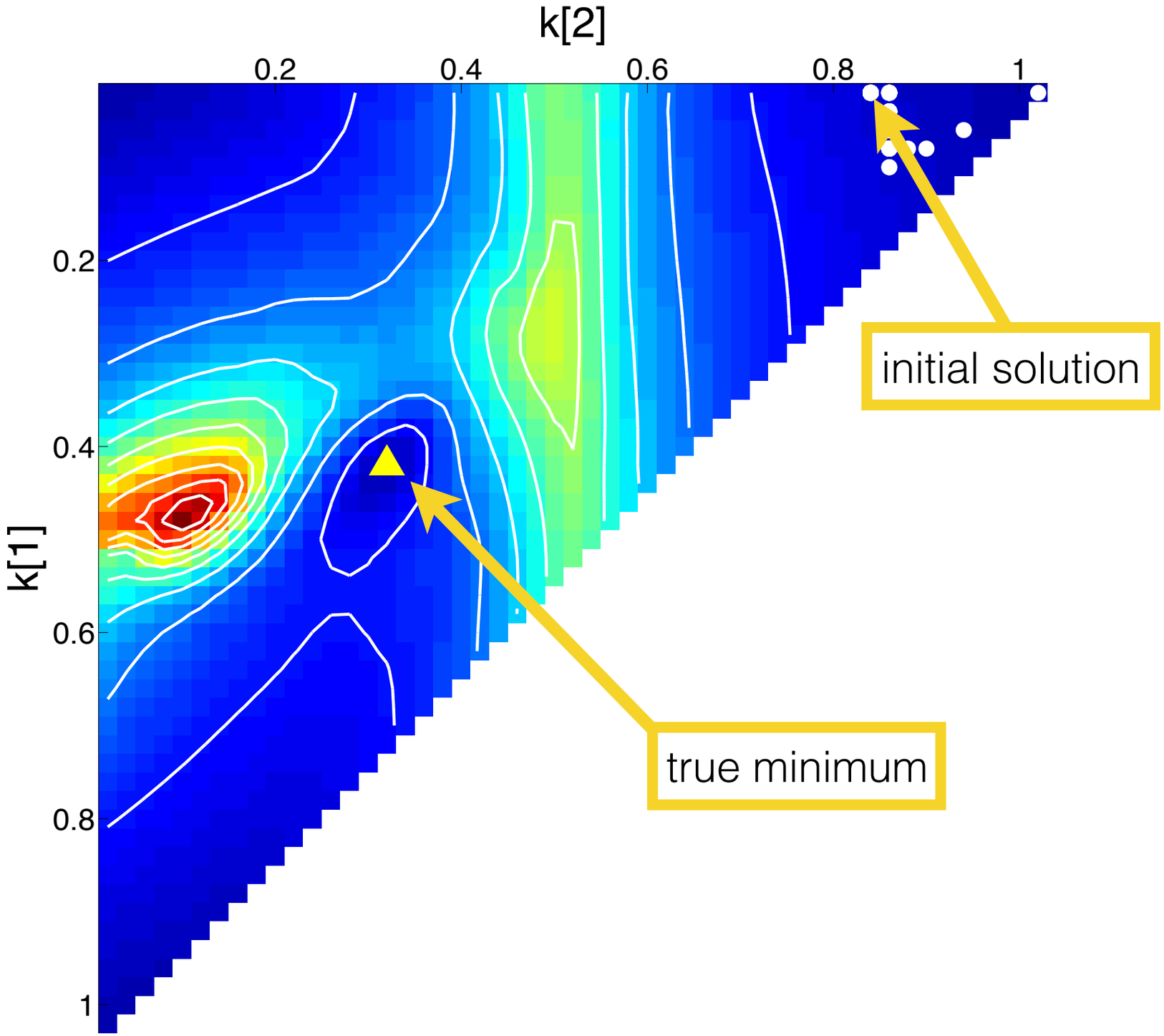
A toy example in 1D



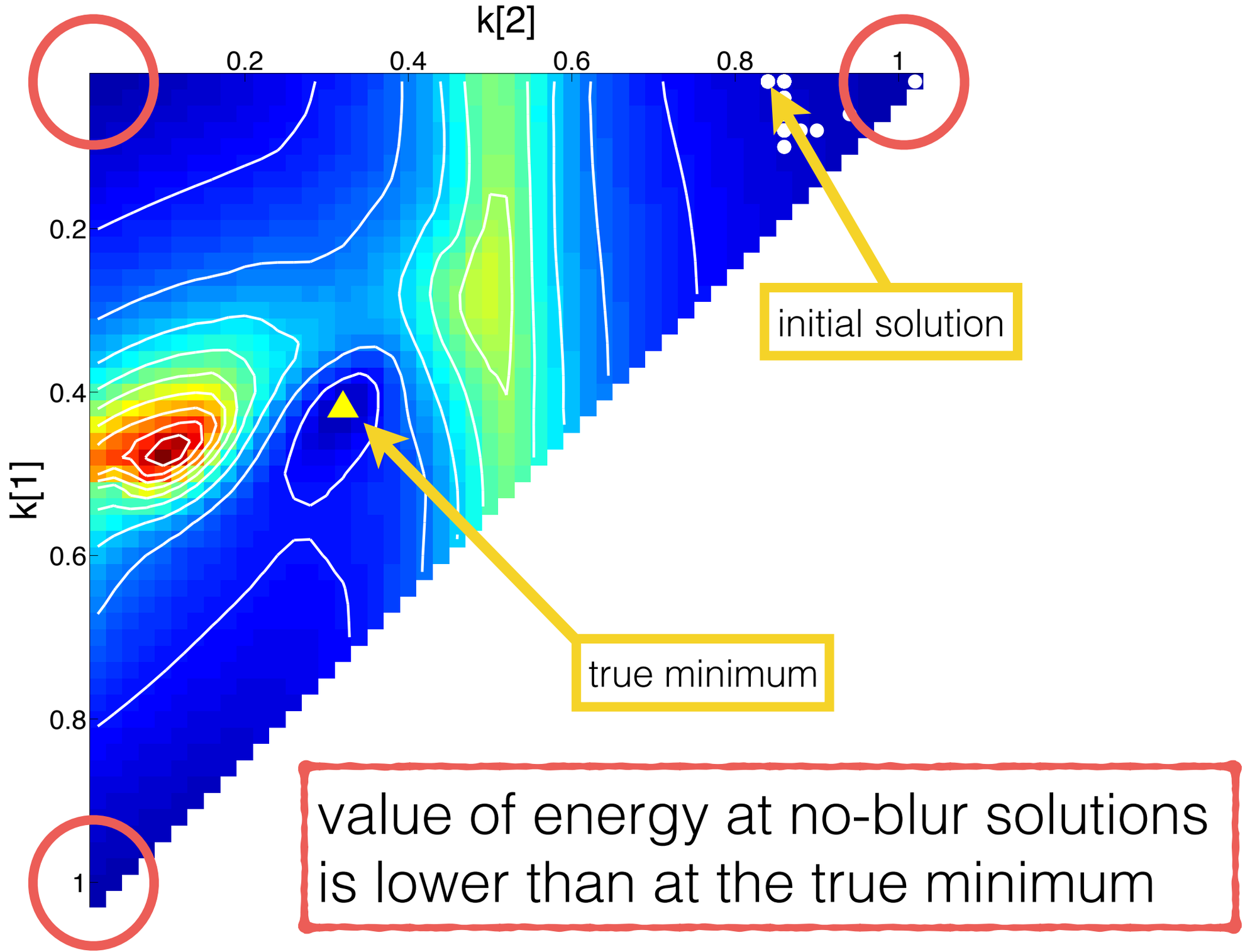
A toy example in 1D



A toy example in 1D



A toy example in 1D



value of energy at no-blur solutions is lower than at the true minimum

Attempt #2: Approximate solution

- Projected alternating minimization (PAM) — implementation of Chang and Wong (1998)

$$u^t \leftarrow u^t - \epsilon \left(k_-^t * (k^t * u^t - f) - \lambda \nabla \cdot \frac{\nabla u^t}{\|\nabla u^t\|_2} \right)$$

$$k^t \leftarrow k^t - \epsilon \left(u_-^t * (k^t * u^t - f) \right)$$

$$k^{t+2/3} \leftarrow \max\{k^{t+1/3}, 0\}, \qquad k^{t+1} \leftarrow \frac{k^{t+2/3}}{\|k^{t+2/3}\|_1}$$

- It works!

Where's Wally?

- What is the difference between AM and PAM that makes PAM work?

- Why does it make it work?

Comparison of AM and PAM

- The first step (image deblurring) is identical

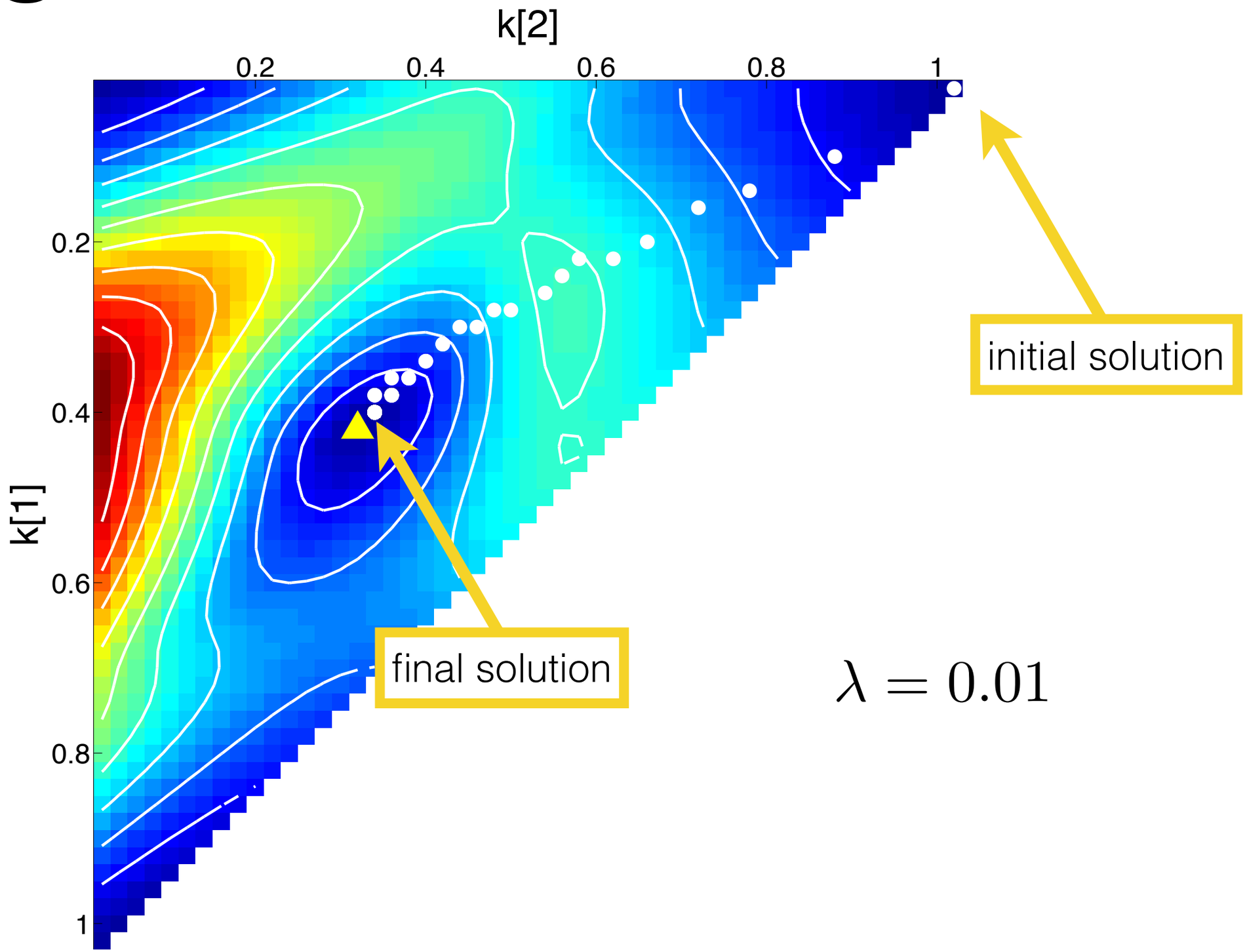
$$u^t \leftarrow u^t - \epsilon \left(k^t_- * (k^t * u^t - f) - \lambda \nabla \cdot \frac{\nabla u^t}{\|\nabla u^t\|_2} \right)$$

- The second step separates the normalization and the positivity constraints from the minimization step

$$k^t \leftarrow k^t - \epsilon \left(u^t_- * (k^t * u^t - f) \right)$$

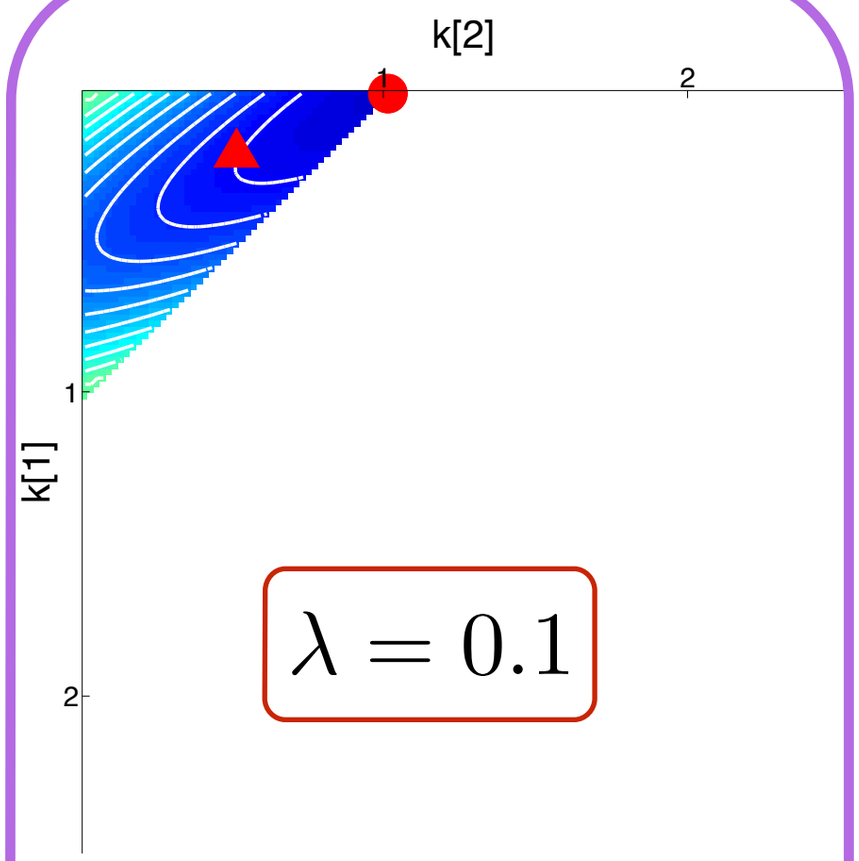
$$k^{t+2/3} \leftarrow \max\{k^{t+1/3}, 0\}, \qquad k^{t+1} \leftarrow \frac{k^{t+2/3}}{\|k^{t+2/3}\|_1}$$

A gradient descent??



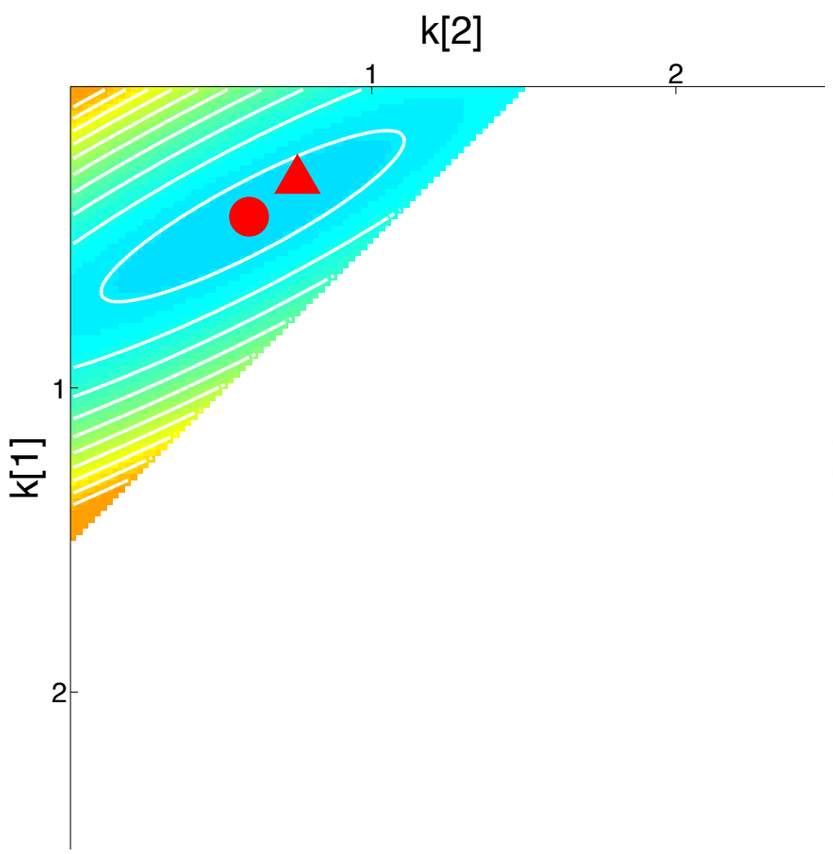
$\lambda = 0.01$

Normalization is the key

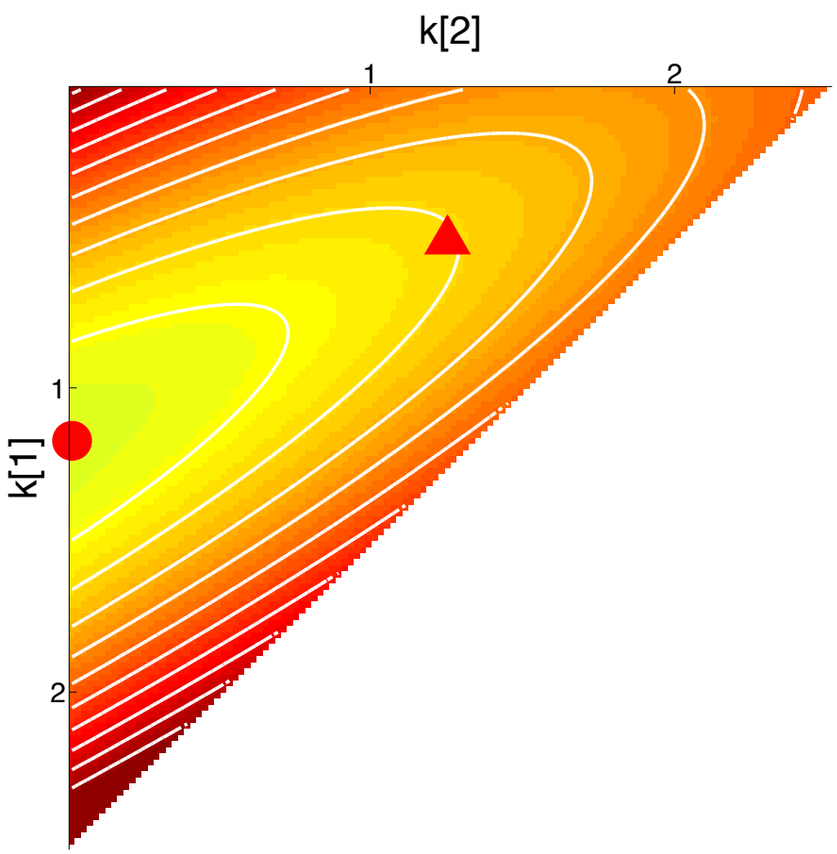


$\lambda = 0.1$

$\|k\|_1 = 1$

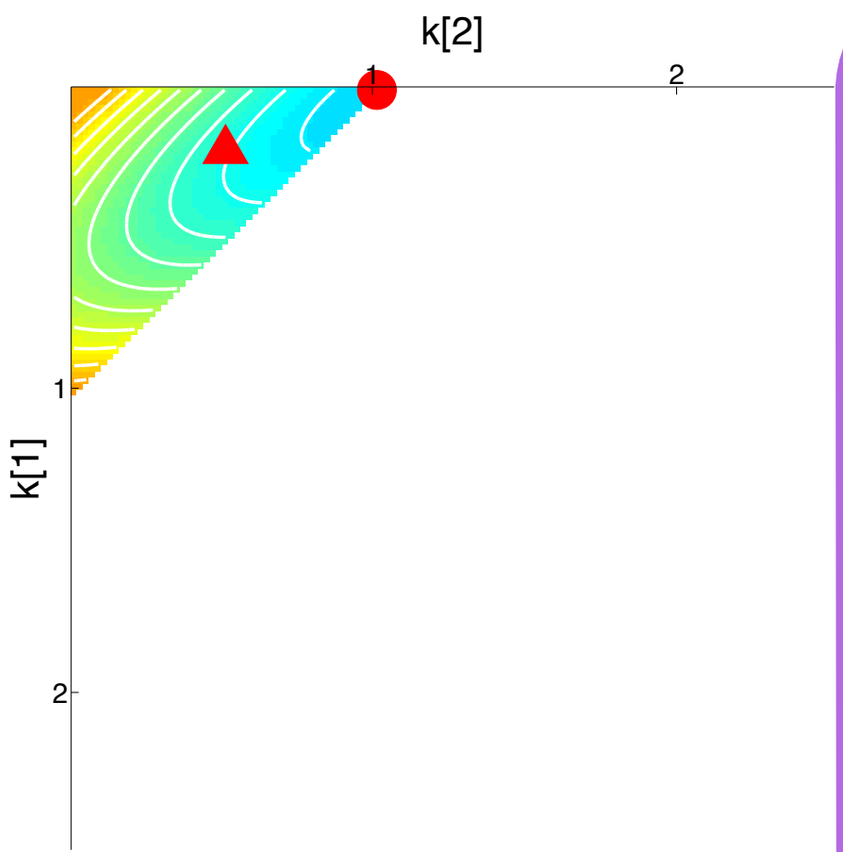


$\|k\|_1 = 1.5$

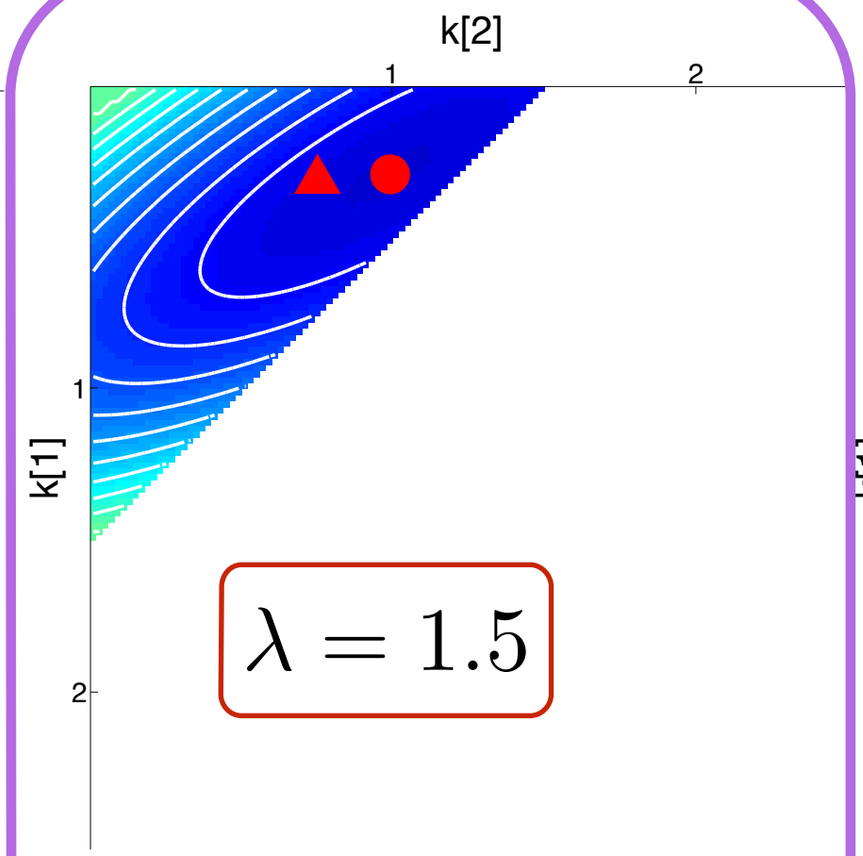


$\|k\|_1 = 2.5$

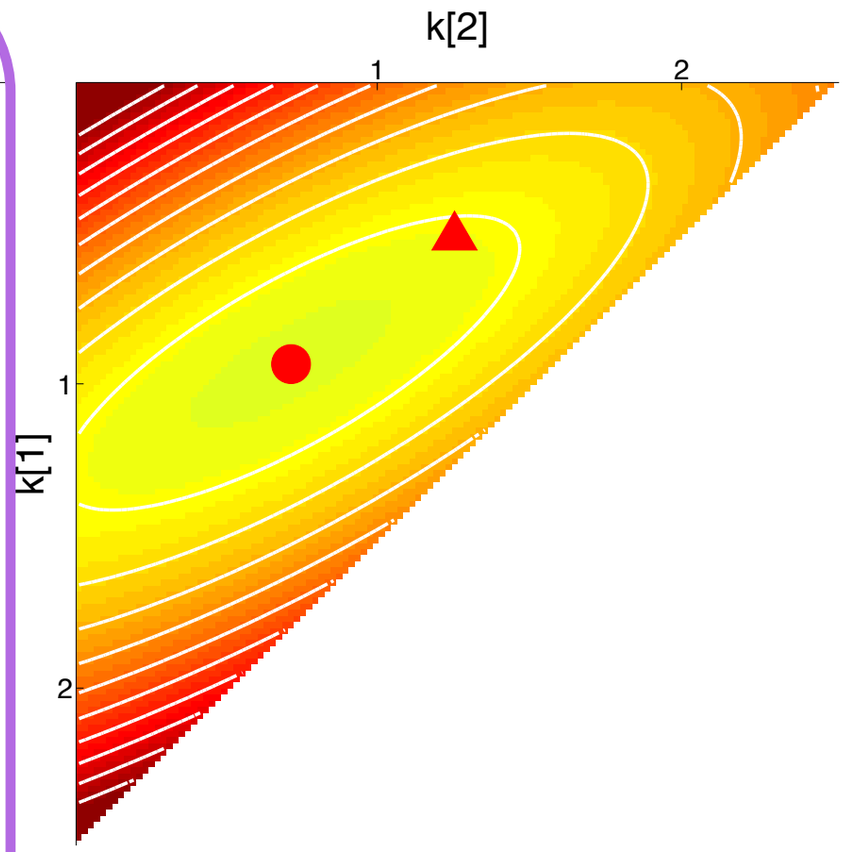
Normalization is the key



$\|k\|_1 = 1$

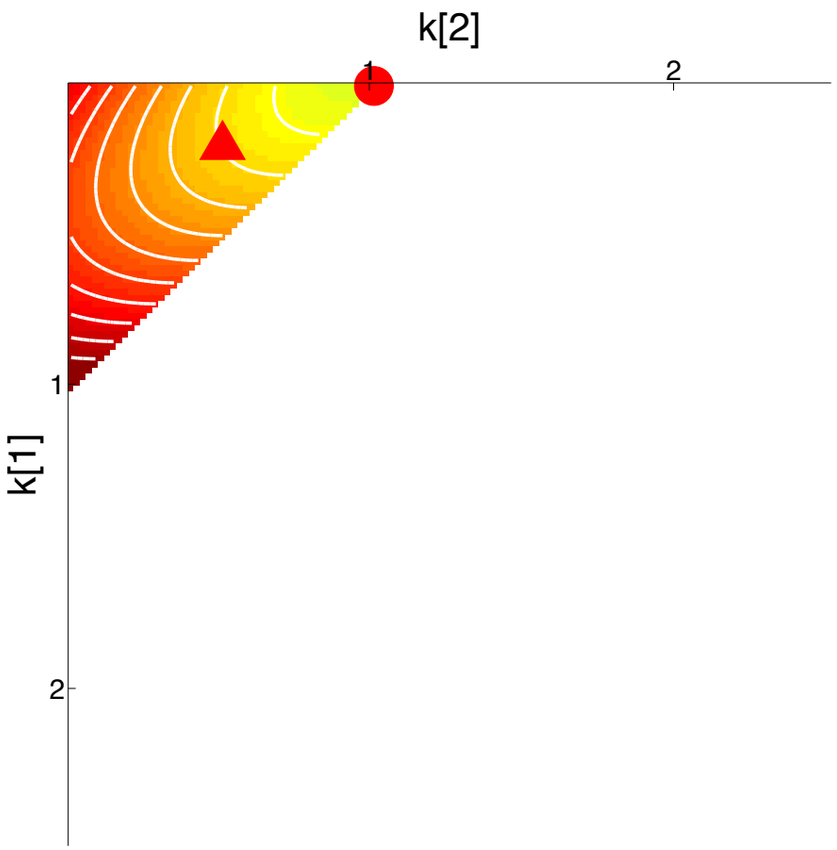


$\|k\|_1 = 1.5$

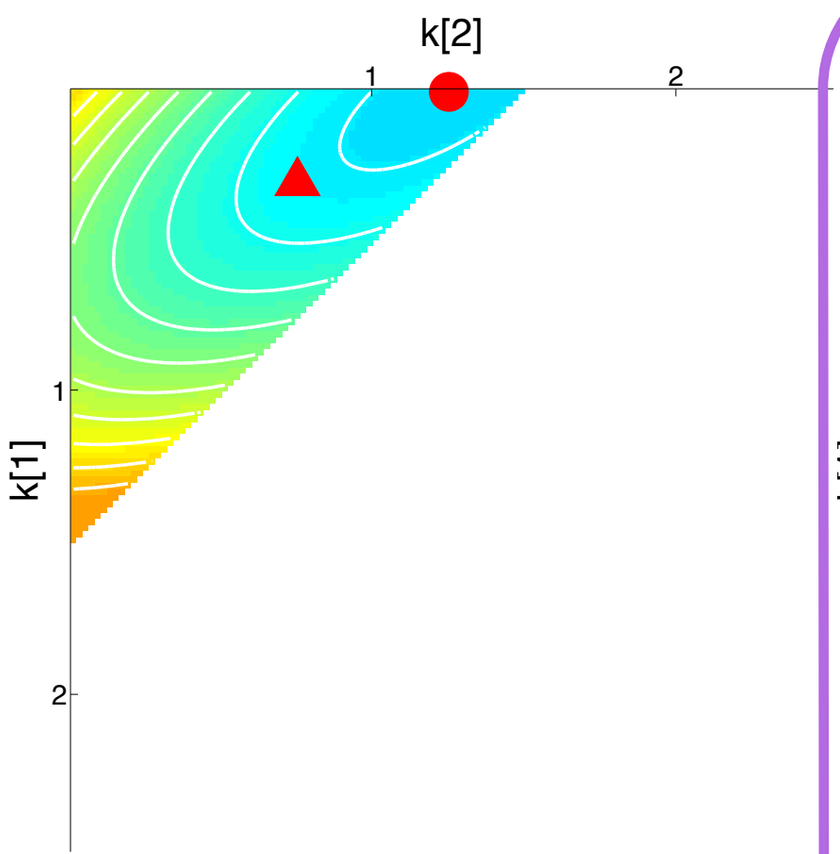


$\|k\|_1 = 2.5$

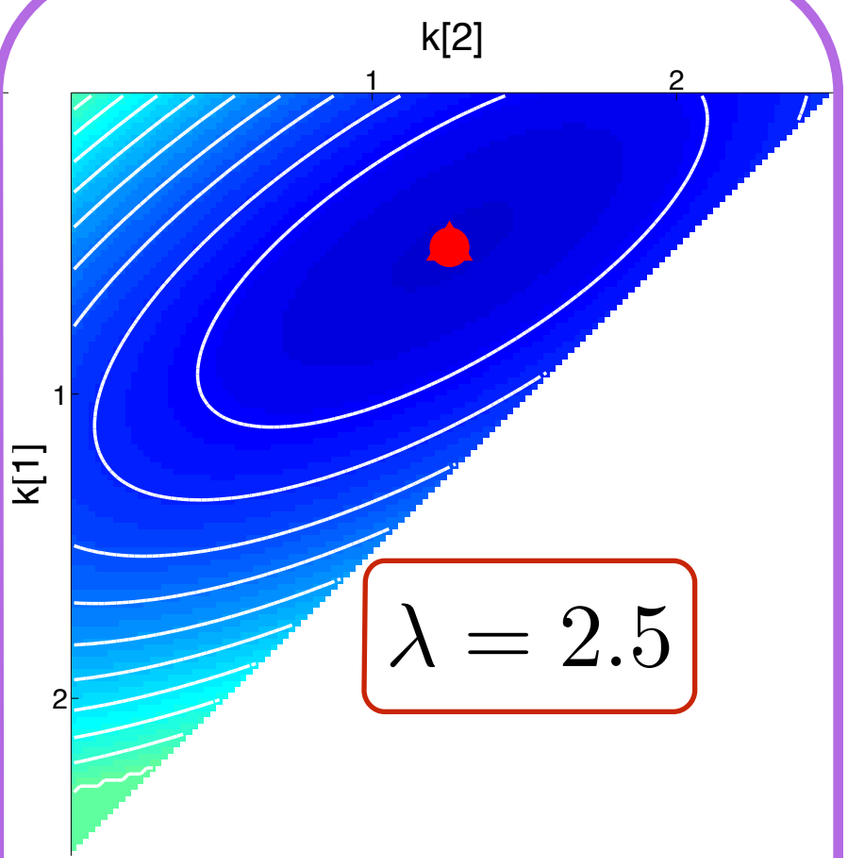
Normalization is the key



$\|k\|_1 = 1$

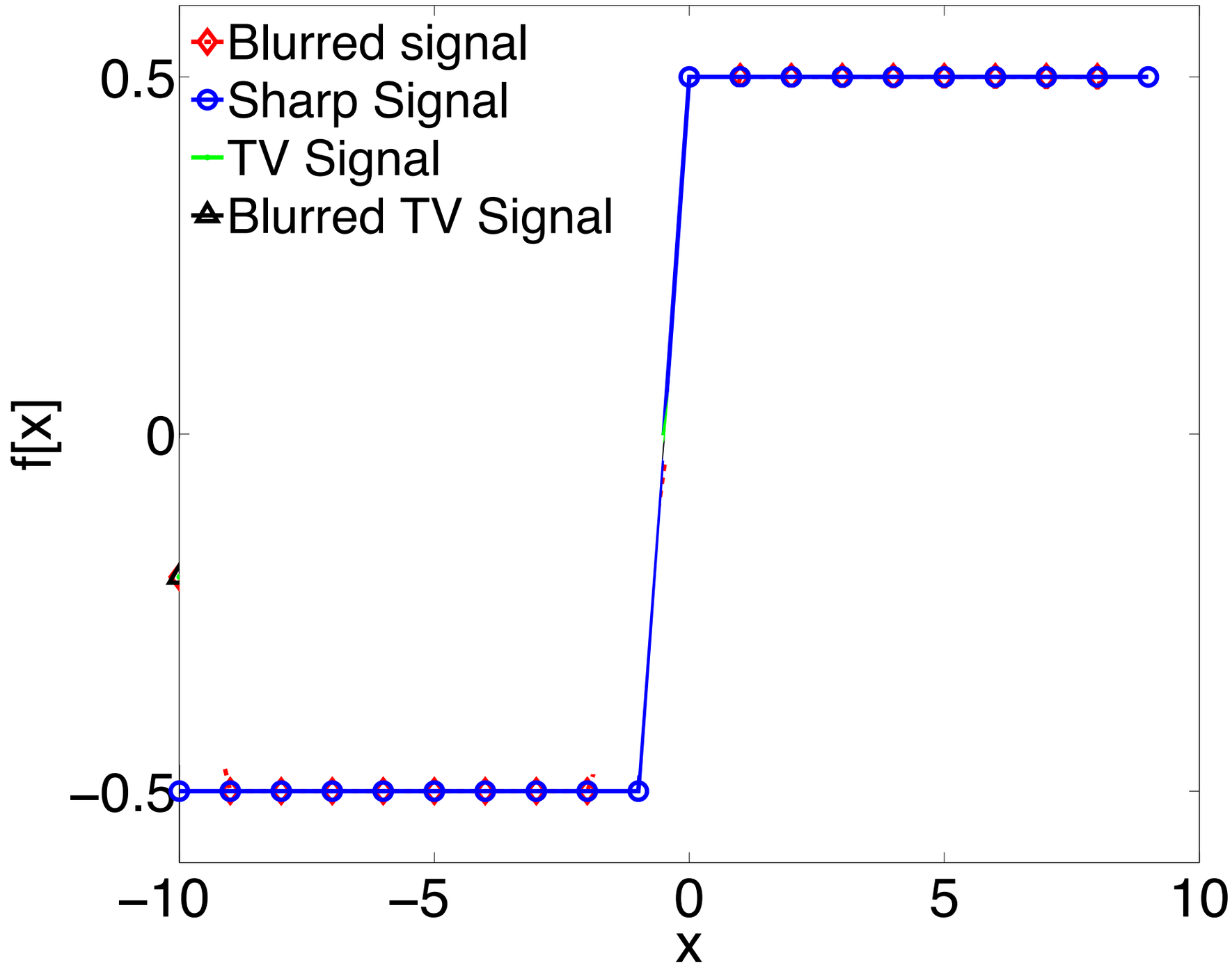


$\|k\|_1 = 1.5$

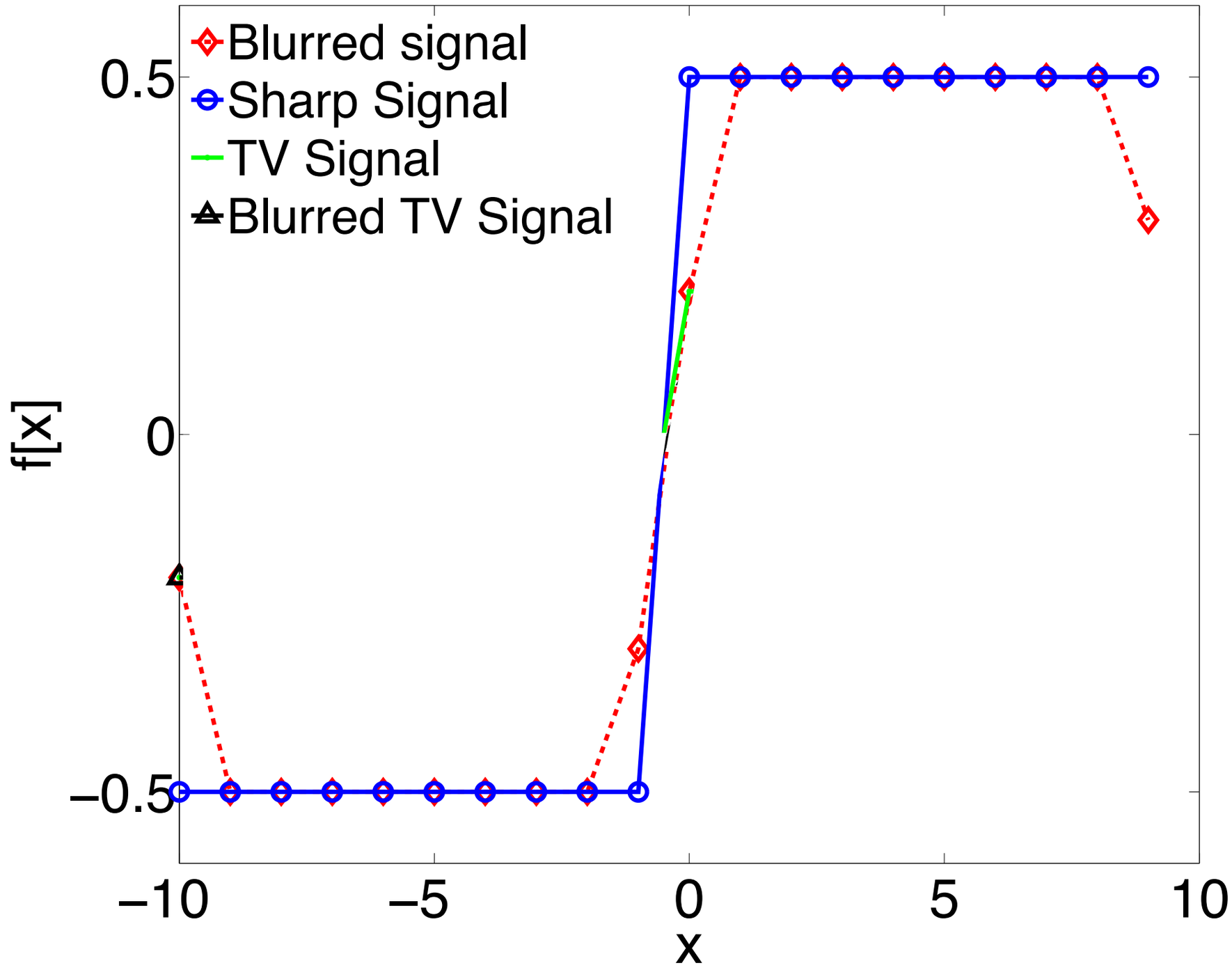


$\|k\|_1 = 2.5$

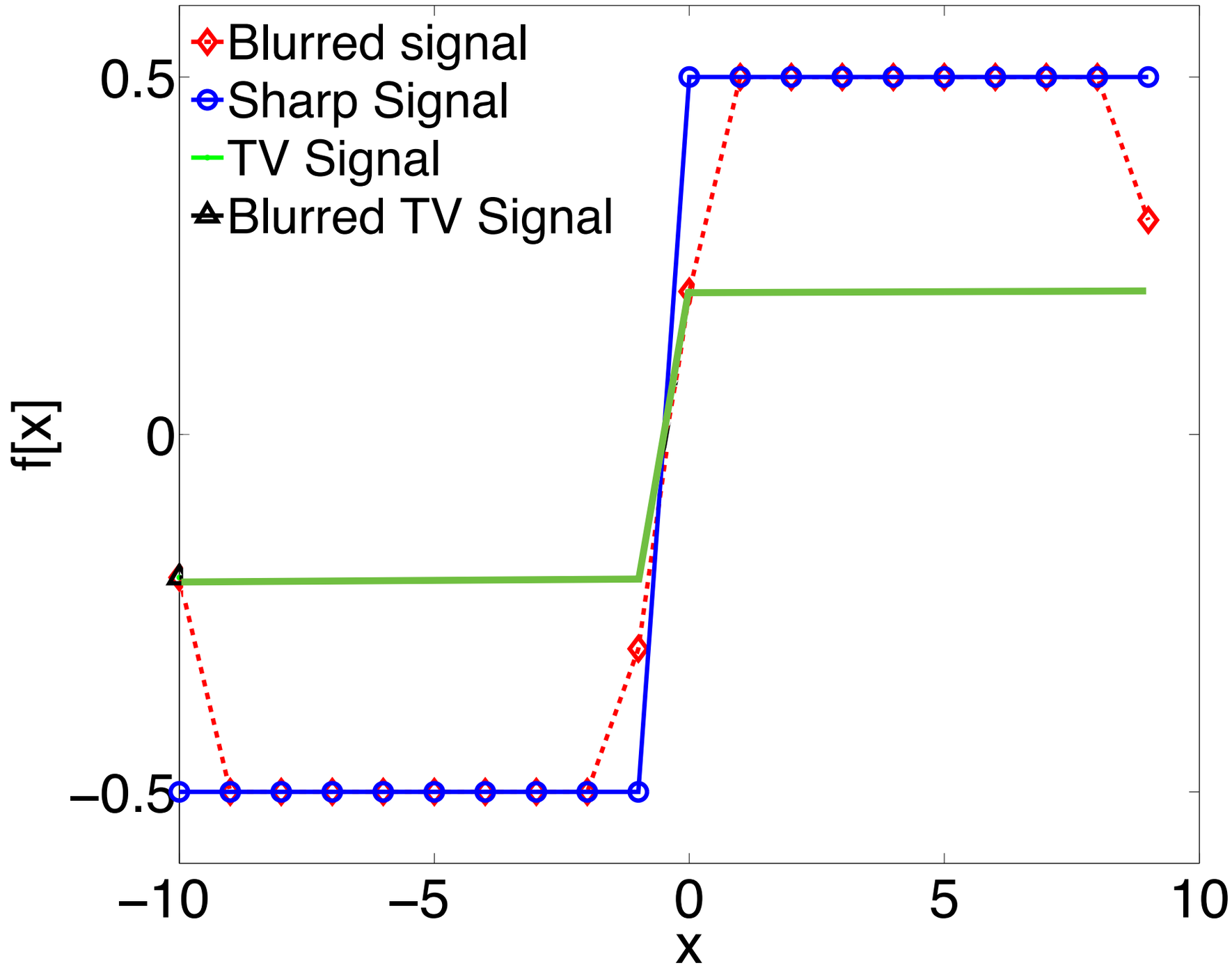
AM on a step function



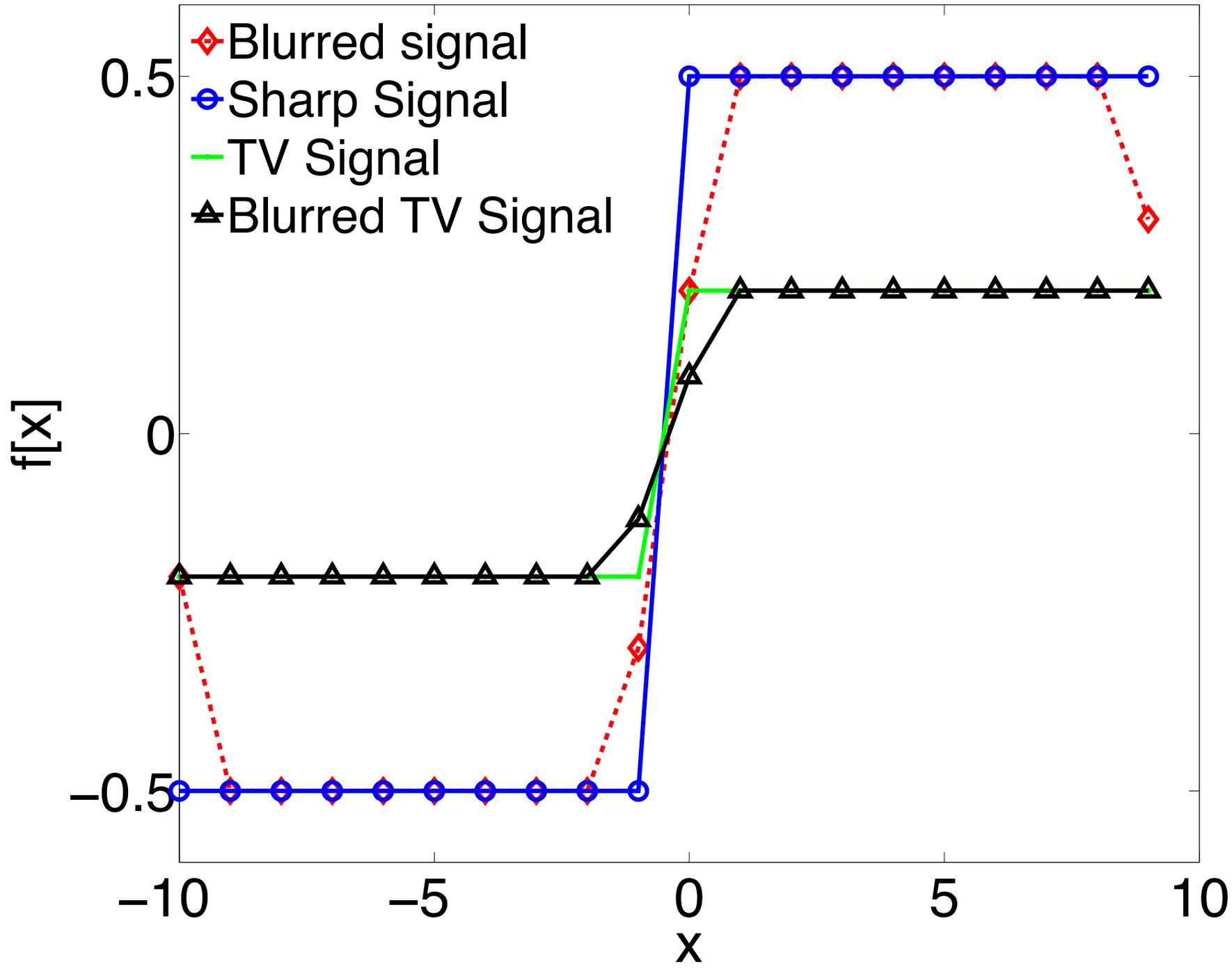
AM on a step function



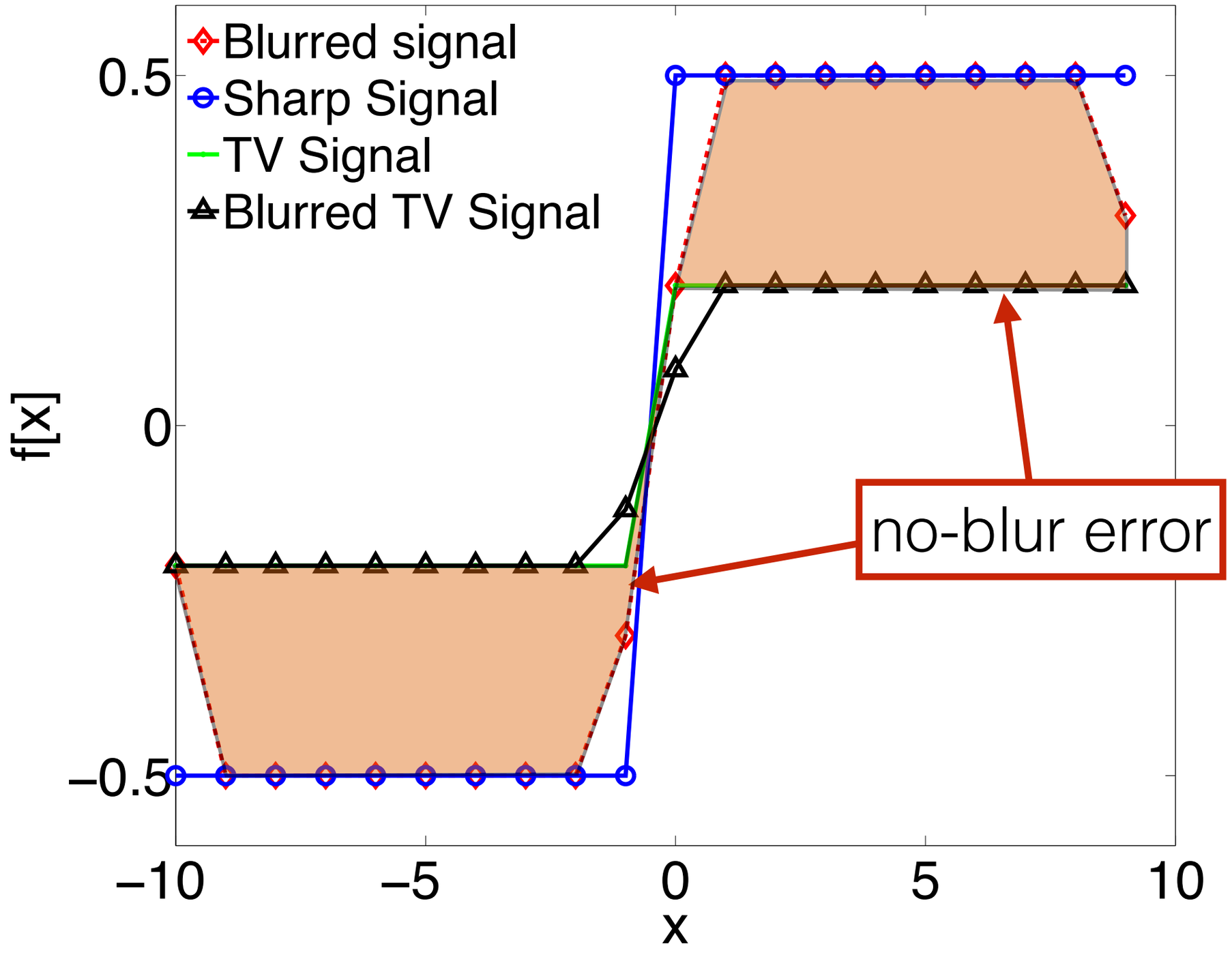
AM on a step function



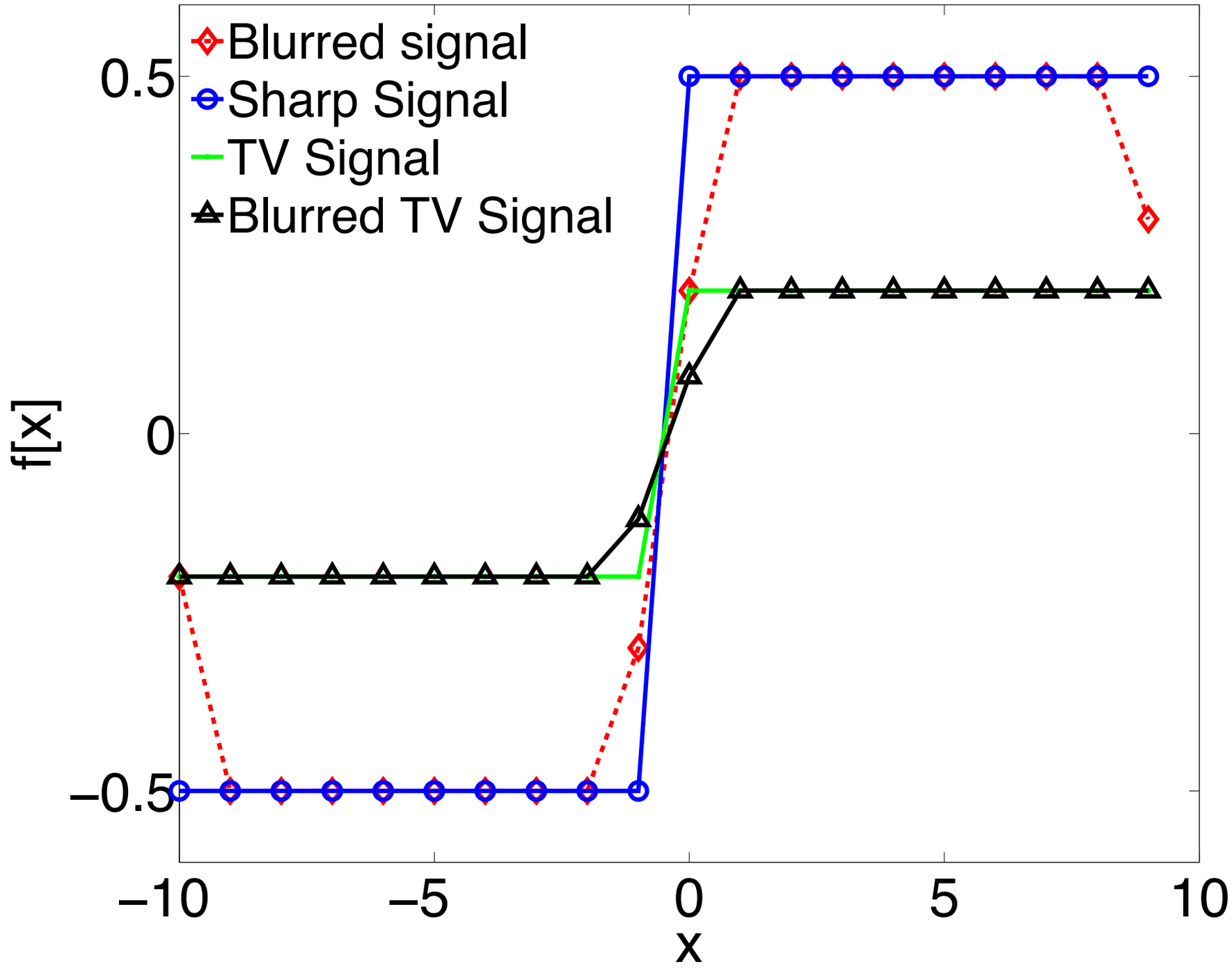
AM on a step function



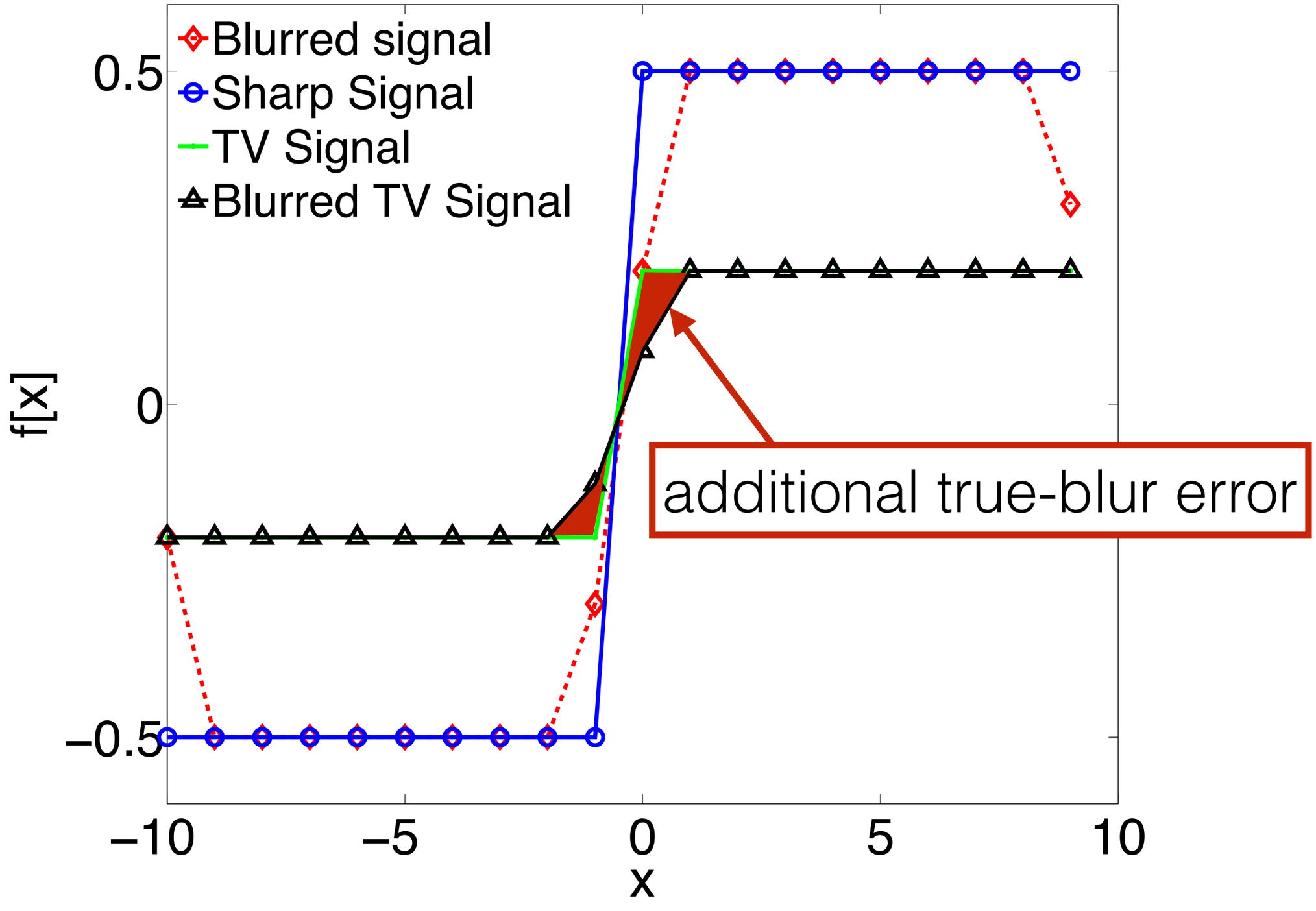
AM on a step function



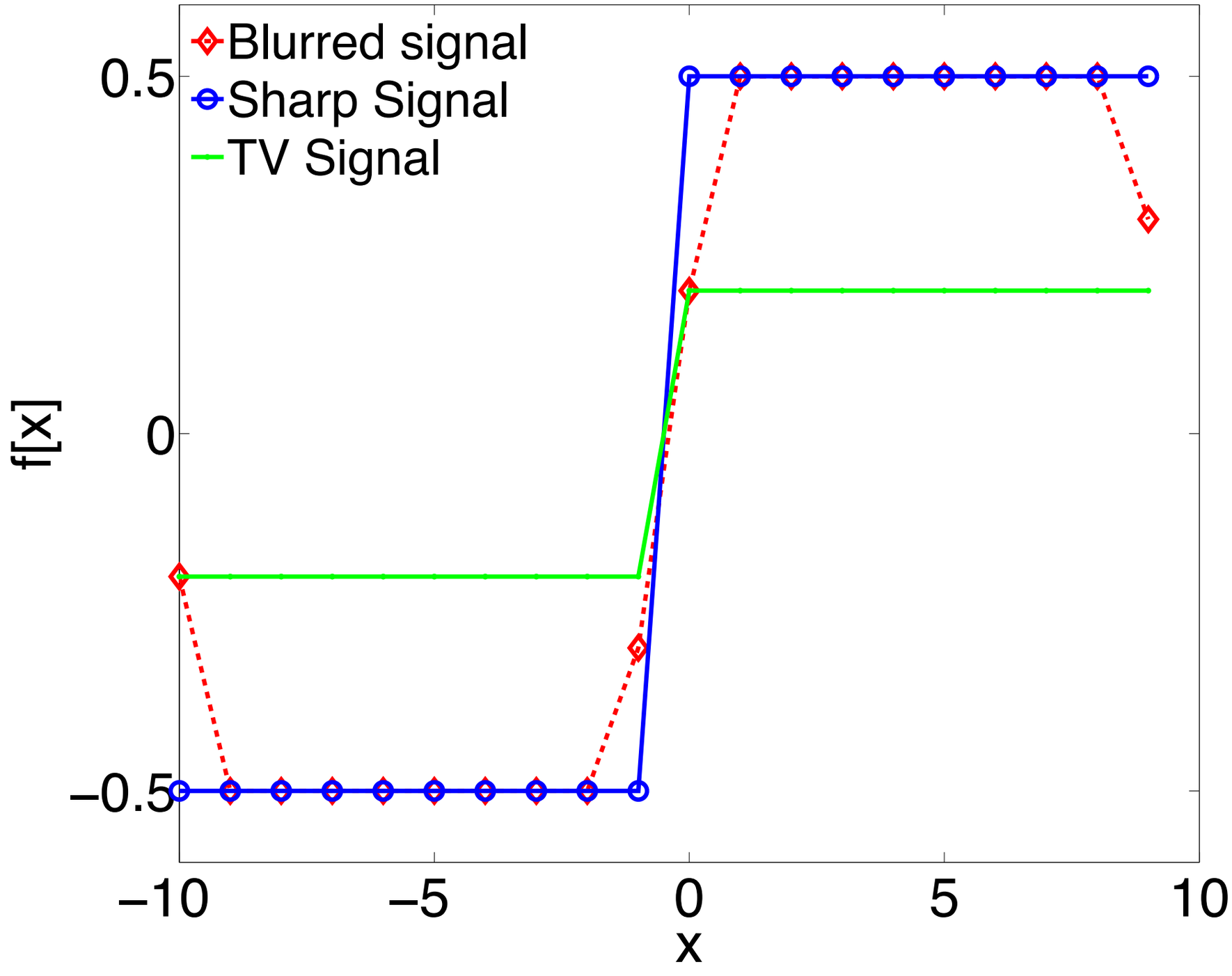
AM on a step function



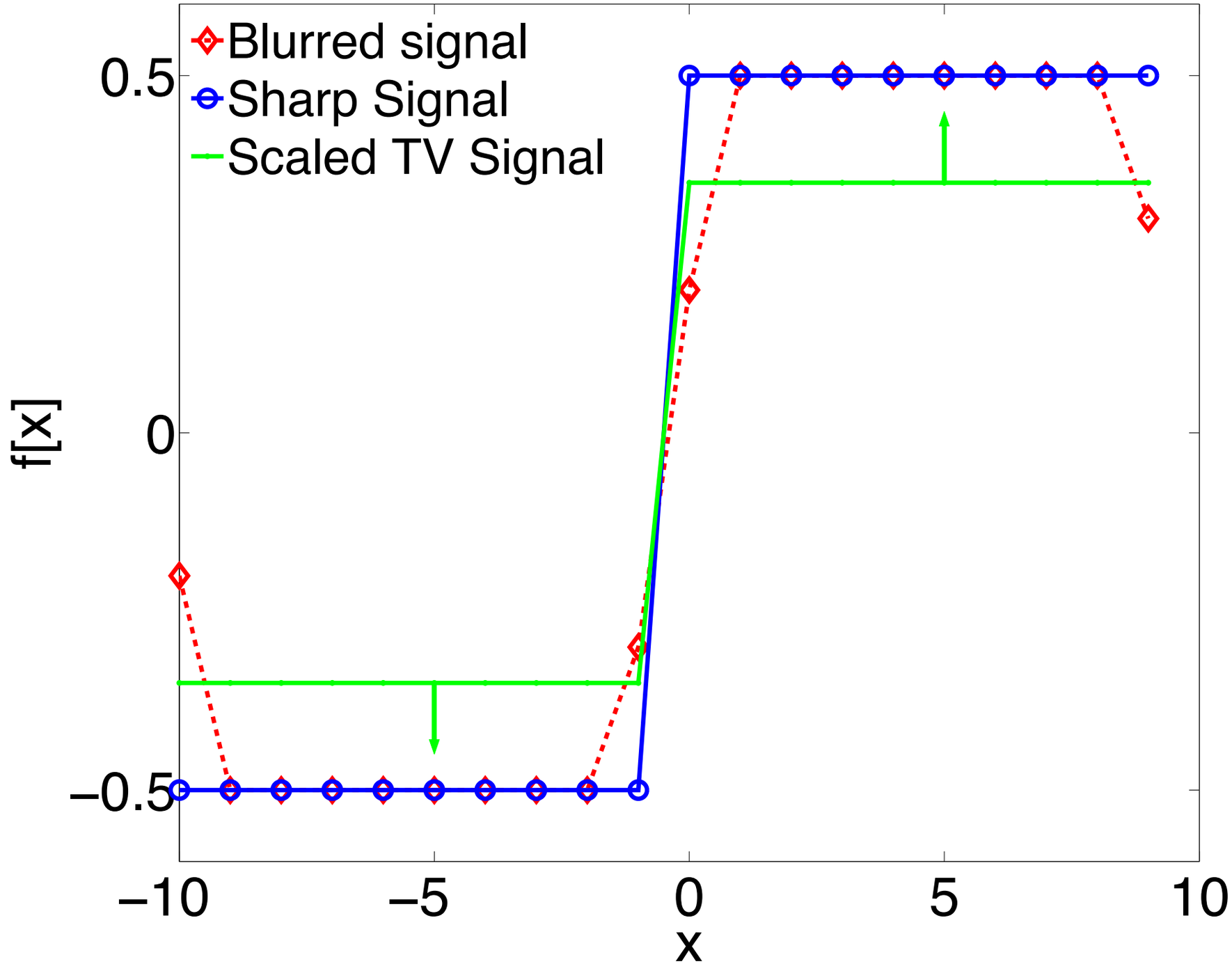
AM on a step function



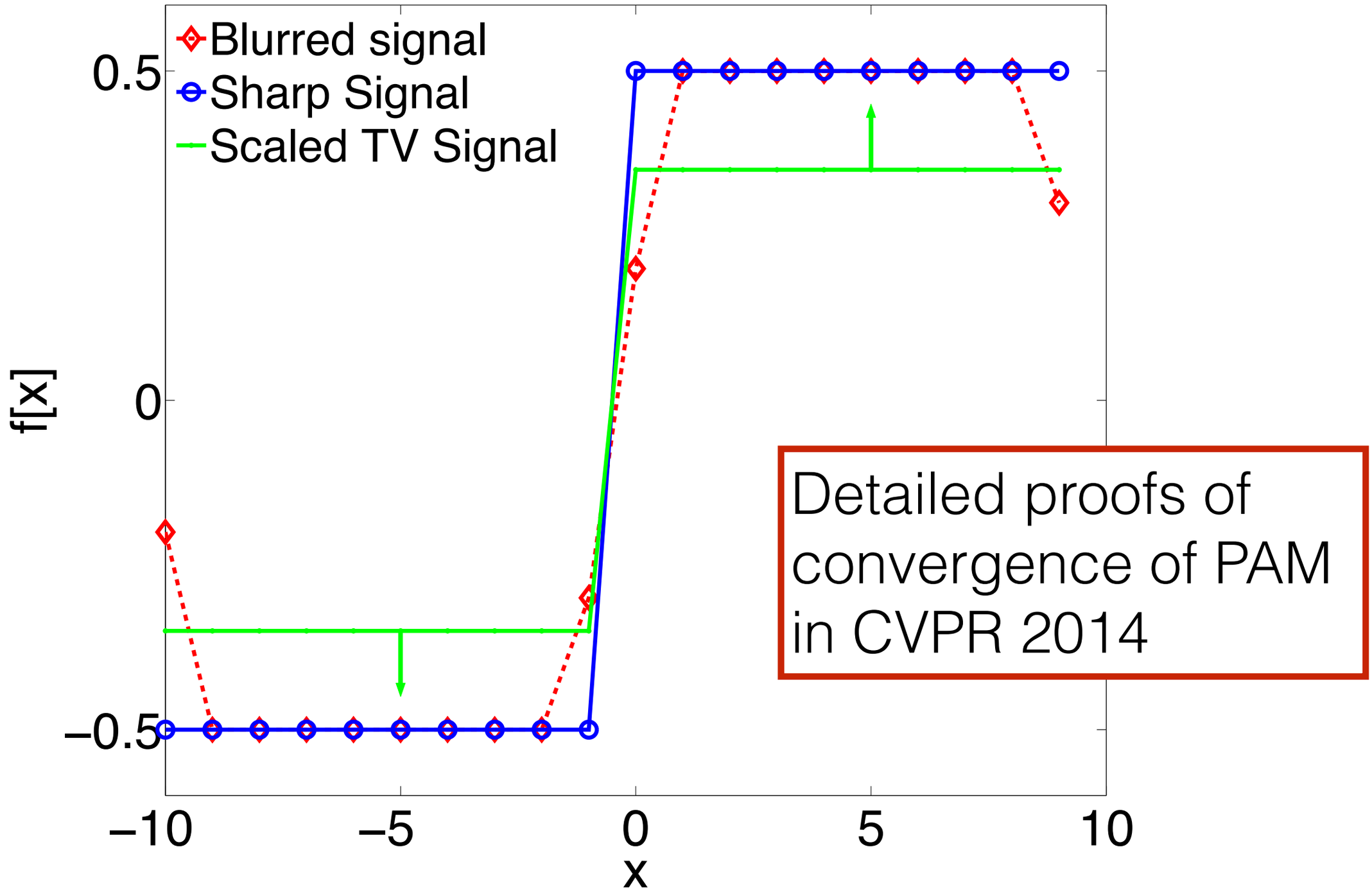
PAM on a step function



PAM on a step function



PAM on a step function



Detailed proofs of convergence of PAM in CVPR 2014

Technical details

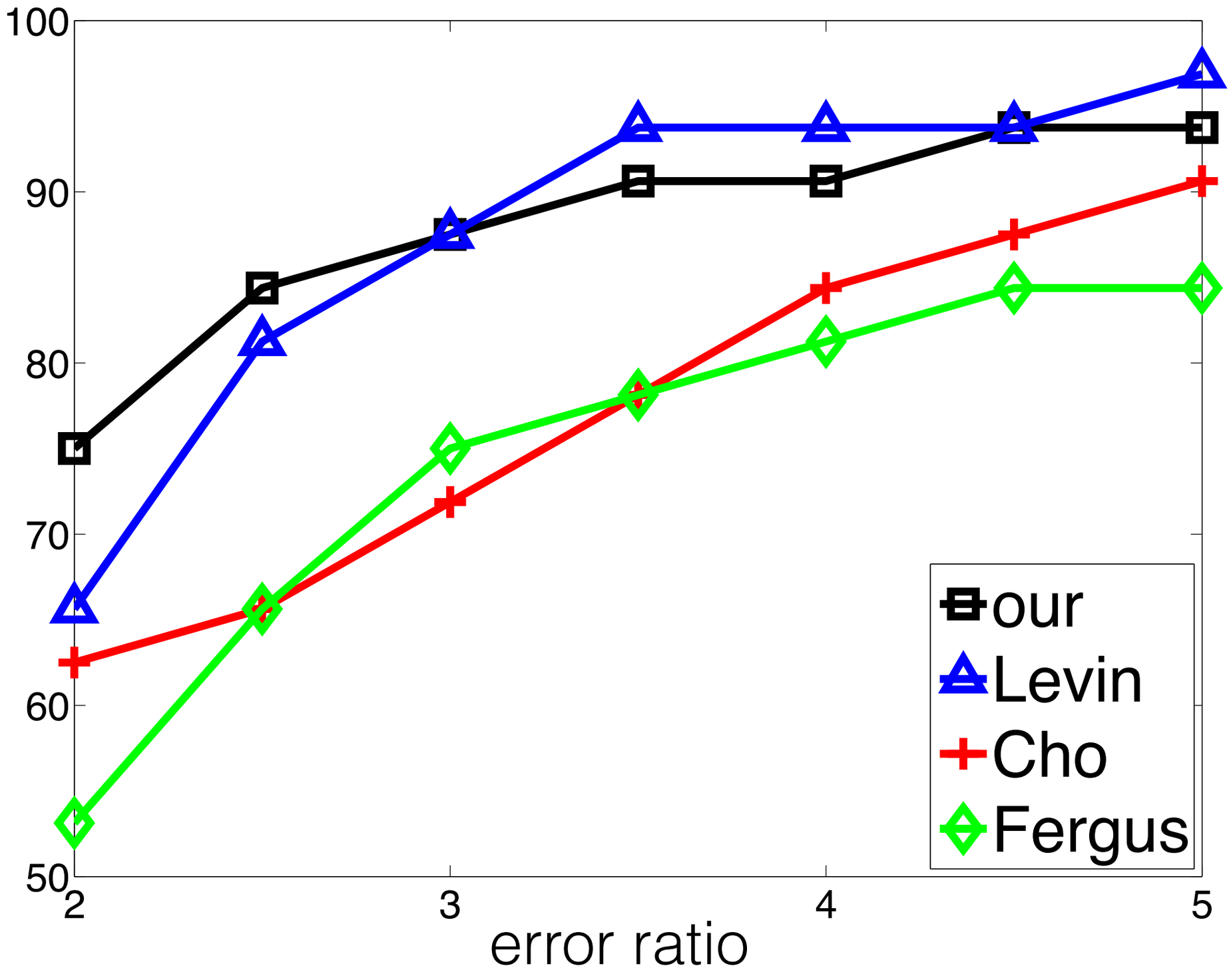
- As in most current implementations we used a pyramid scheme
- Adaptation of the regularization parameter is needed
- Boundary conditions: None as we use the exact blur model

The PAM algorithm

Data: f, size of blur, initial large λ, final λ_{min}
Result: u,k

$u^0 \leftarrow \text{pad}(f)$;
$k^0 \leftarrow \text{uniform}$;
while *not converged* **do**
  $u^{t+1} \leftarrow u^t - \epsilon_u \left(k_-^t \bullet (k^t \circ u^t - f) - \lambda \nabla \cdot \frac{\nabla u^t}{|\nabla u^t|} \right)$;
  $k^{t+1/3} \leftarrow k^t - \epsilon_k \left(u_-^{t+1} \circ (k^t \circ u^{t+1} - f) \right)$;
  $k^{t+2/3} \leftarrow \max\{k^{t+1/3}, 0\}$;
  $k^{t+1} \leftarrow \frac{k^{t+2/3}}{\|k^{t+2/3}\|_1}$;
  $\lambda \leftarrow \max\{0.99\lambda, \lambda_{min}\}$;
  $t \leftarrow t + 1$;
end
$u \leftarrow u^{t+1}$;
$k \leftarrow k^{t+1}$;

Experiments

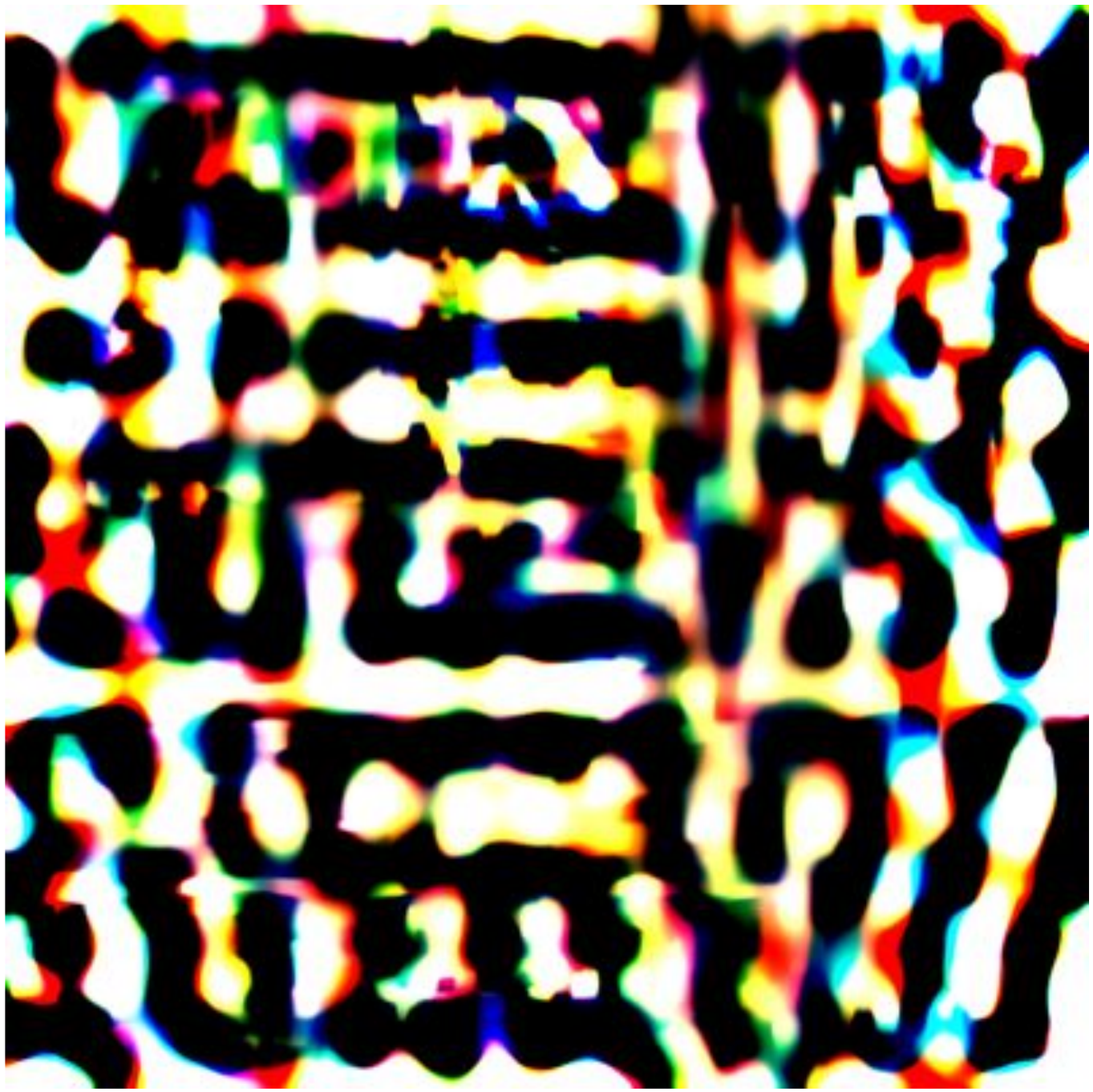


Blurry image

Cho and Lee (2009)

Fergus et al (2006)

Hirsch et al (2011)



Shan et al (2008)

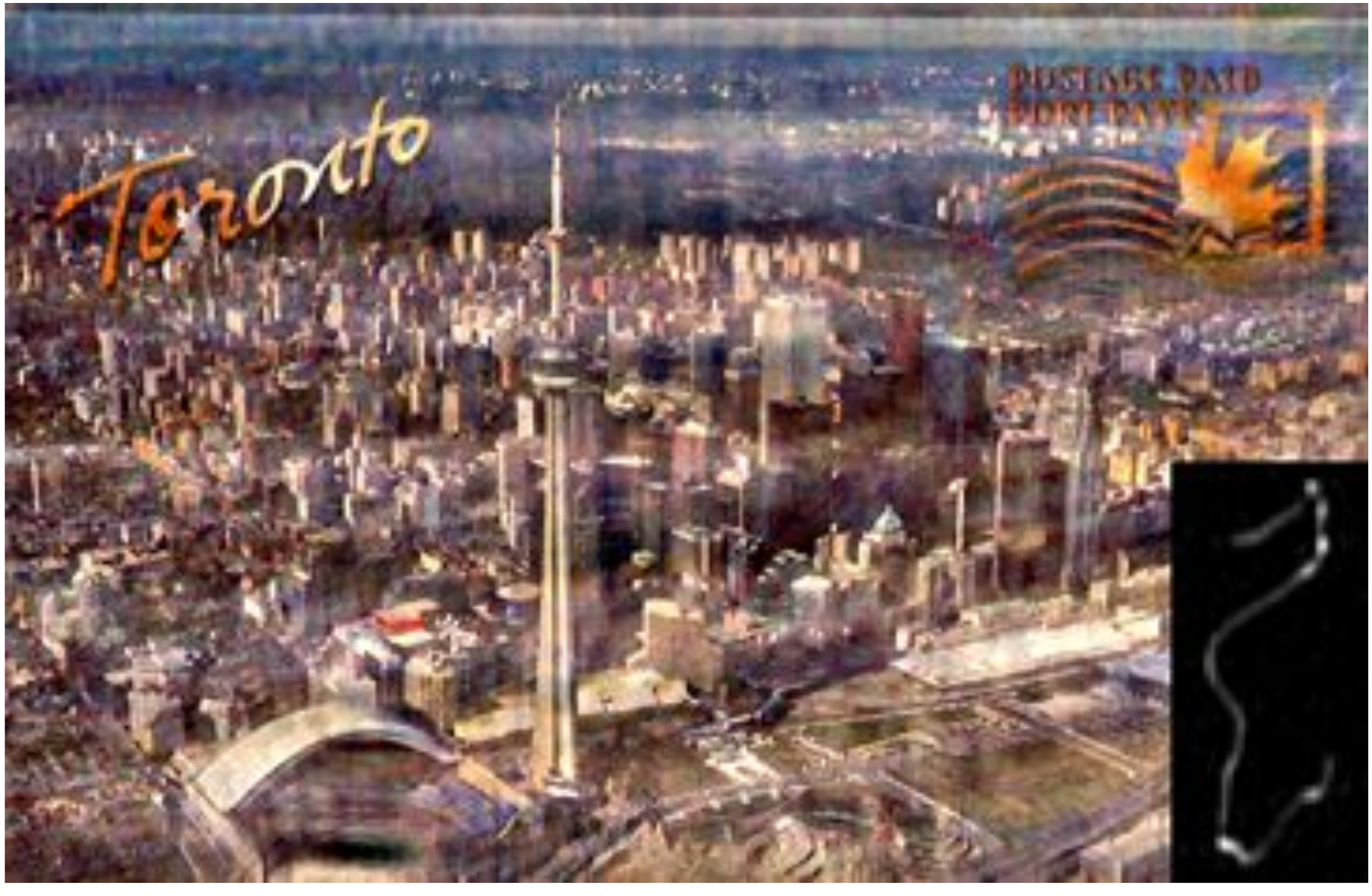
Whyte et al (2011)

Xu and Jia (2010)

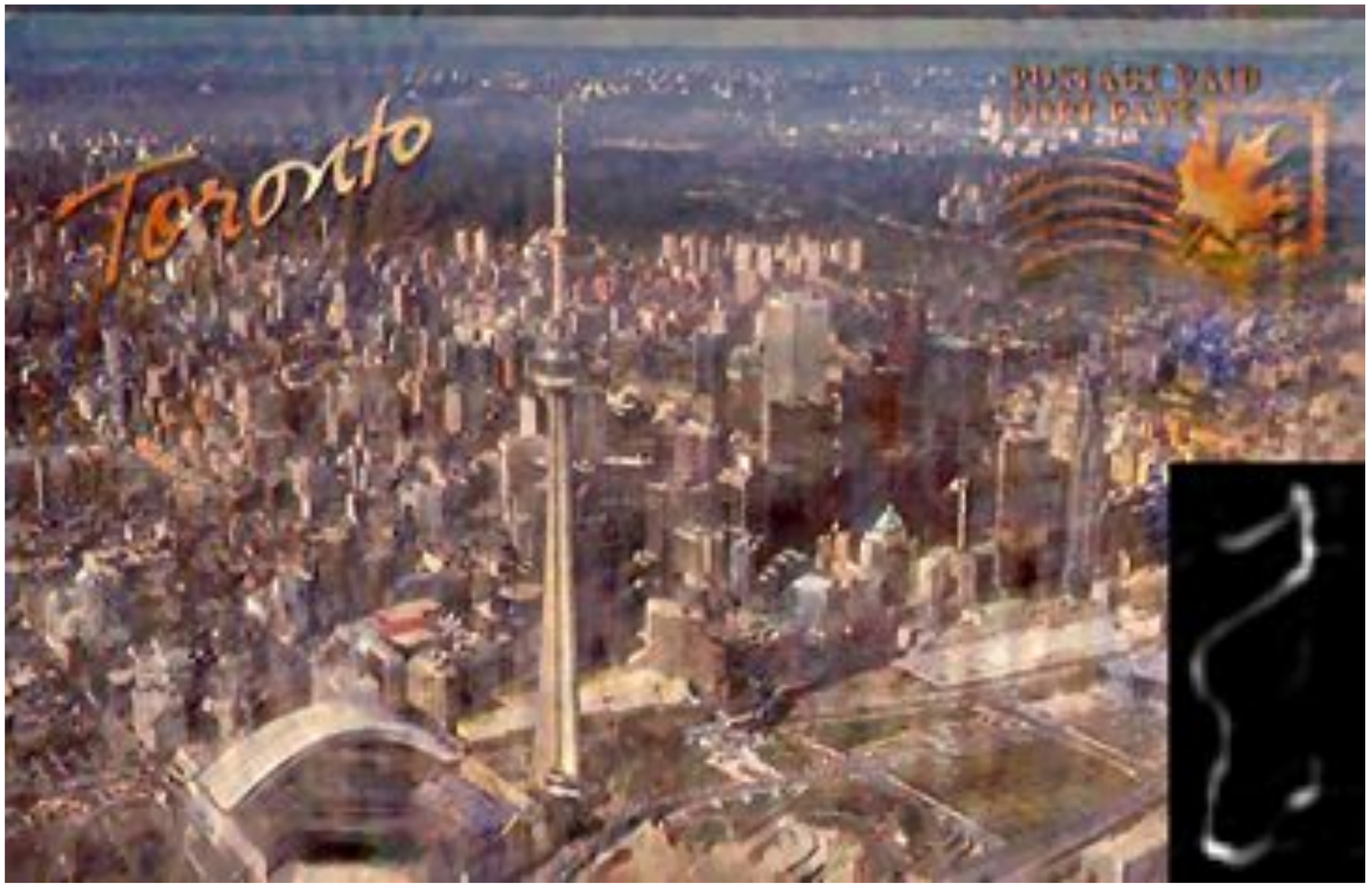
Our (PAM)

Blurred

Xu and Jia (2010)

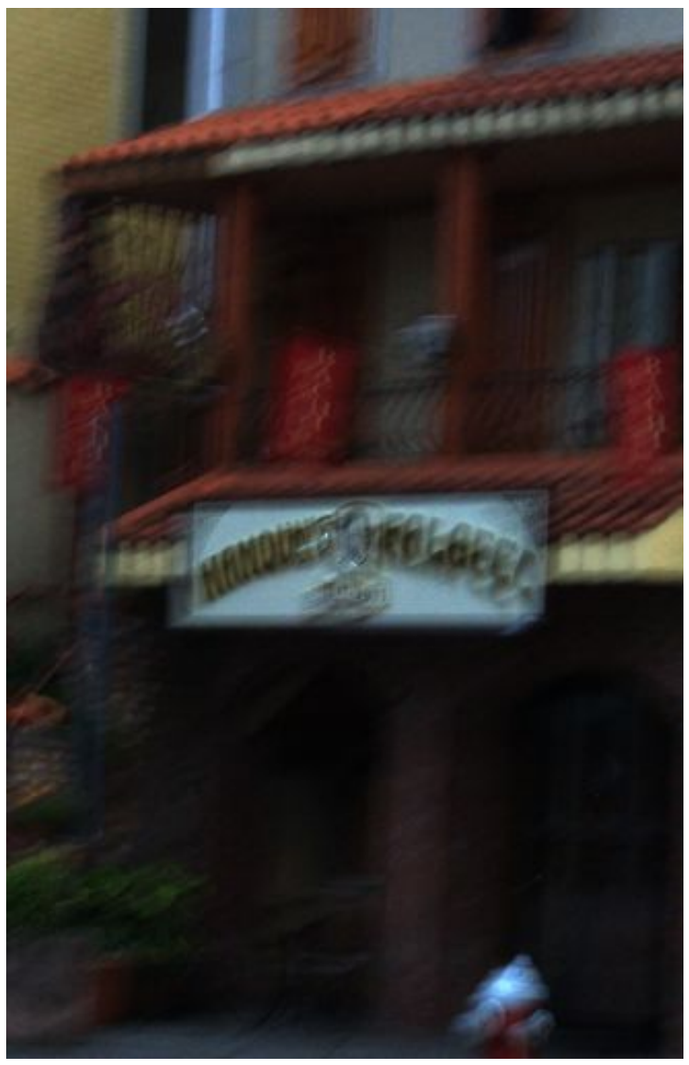


Our (PAM)



One more example

blurry

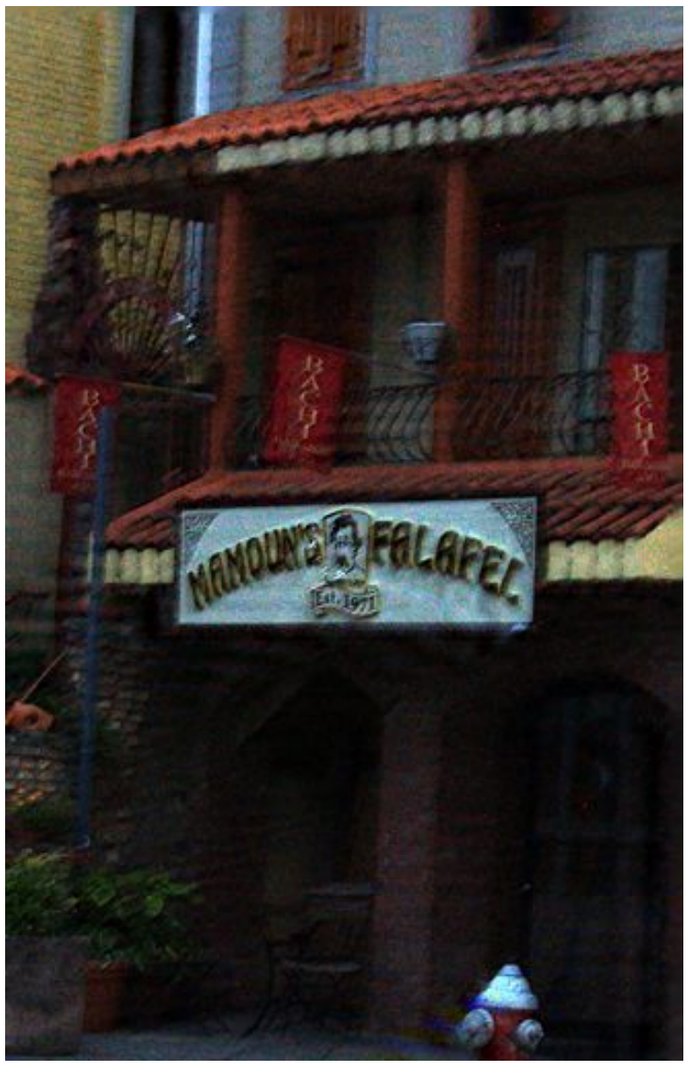


Cho and Lee (2009)

Goldstein and Fattal (2012)

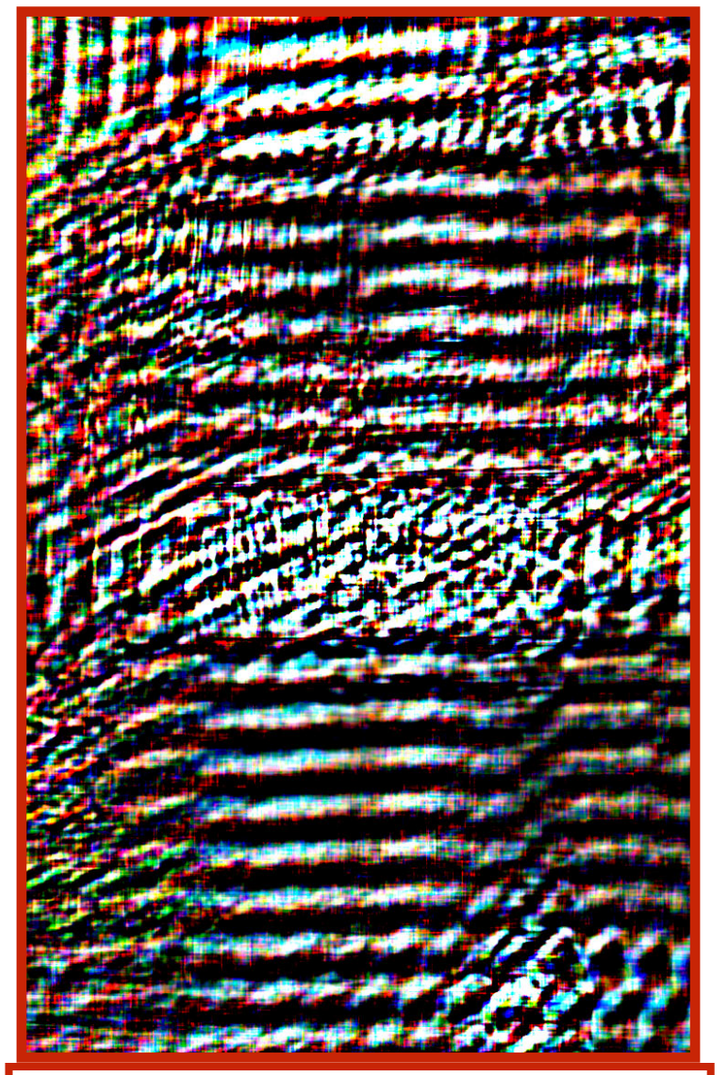
One more example

our (PAM)



Zhong et al (2013)

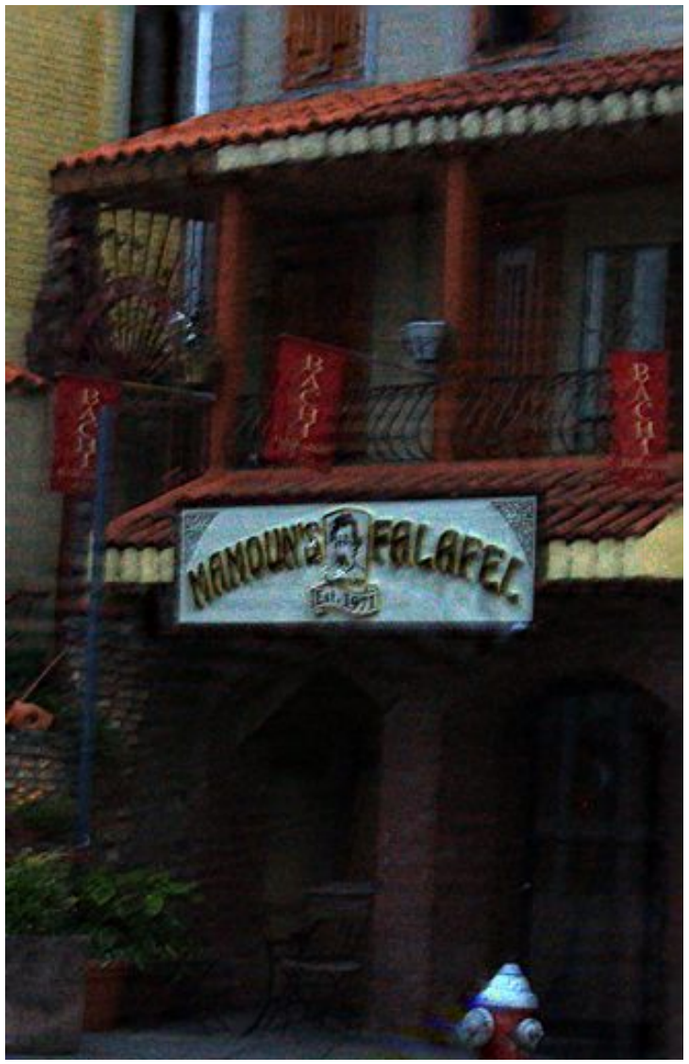
Levin et al (2011)



be wary of the results of others!

One more example

our (PAM)



Zhong et al (2013)

Levin et al (2011)

Conclusions

- We have shown (with theory and experiments) why many alternating minimization algorithms work
  - The reason lies in the normalization of blur (scaling) + regularization parameter
- This 1998 algorithm competes very well with recent more sophisticated algorithms
- Perhaps we should rethink our formulation of blind deconvolution?