# Observations Concerning Reconstructions with Local Support

Per-Erik Forssén

Computer Vision Laboratory, Department of Electrical Engineering

Linköping University, SE-581 83 Linköping, Sweden

April 29, 2002

# 1  Abstract

This report describes how the choice of kernel affects a non-parametric density estimation. Methods for accurate localisation of peaks in the estimated densities are developed for Gaussian and $\cos^2$ kernels. The accuracy and robustness of the peak localisation methods are studied with respect to noise, number of samples, and interference between peaks. Although the peak localisation is formulated in the framework of non-parametric density estimation, the results are also applicable to associative learning with localised responses.

# 2  Introduction

Assume that we have a set of measurements, $s_n$, which are drawn from a discrete distribution contaminated with additive noise. The measurements are known to only assume values in a limited range of the real axis i.e. $s_n \in [L_1, L_2]$. Such a distribution will have a *probability density function* (PDF) that looks something like figure 1. We now want to estimate the distribution of the discrete variable in a robust and accurate manner. As we shall see, an effective way to perform this estimation is to determine the locations of the peaks in the distribution of $s_n$.
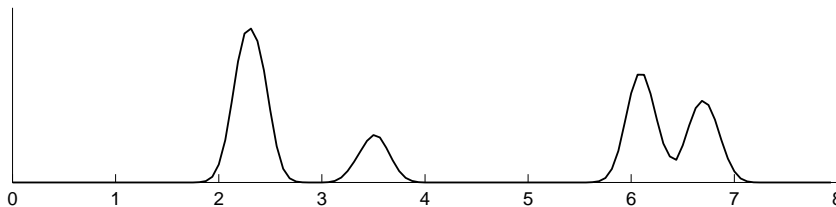


Figure 1: Example of PDF.

1

Since we have no a priori knowledge of the distribution of the measurements, we make a *non-parametric density estimation*, commonly referred to as a histogram computation. When computing a histogram, we want to find a representation of the PDF as a set of *bin values*, $\{h_k\}_1^K$. Each bin value is the sum of a *kernel function*, $H_k(s_n)$, applied on the samples.

$$h_k = \sum_{n=1}^{N} H_k(s_n) \quad \text{for each} \quad k = 1 \ldots K \tag{1}$$

For ease of computation the kernel function is often chosen as a rectangular function

$$H_k(s_n) = \begin{cases} 1 & \text{when} \quad |s_n - c_k| < \dfrac{L_2 - L_1}{2(K-1)} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Where $\{c_k\}_1^K$ is the set of bin centres, usually uniformly distributed across the range, $[L_1, L_2]$, of the variable under study ($s_n$ in our case).

The kernel function in equation 2 is undesirable, since bin value computation for a single sample lacks a unique inverse (it is not *injective*), thus the reconstruction of peak locations can not be very accurate when the number of samples, $N$, is low.

If the kernel function is known, we can make use of this knowledge in the estimation of peak locations. This is evident from the expectation value of the bin values. For samples $s_n \in F$ with the associated PDF $f(x)$ the expected value of bin $k$ is:

$$\mathrm{E}\left\{\frac{h_k}{N}\right\} = \mathrm{E}\left\{H_k(x)\right\} = \int H_k(x)f(x)dx \tag{3}$$

According to the definition of expectation value for a function of a variate, see for instance [4]. This expression can be rewritten as a convolution:

$$\mathrm{E}\left\{H_k(x)\right\} = \int H(c_k - x)f(x)dx = (H * f)(c_k) \tag{4}$$

where $H(x)$ is a mirrored[1] kernel function centred around $x = 0$, and $c_k$ is the current bin centre. If we view the bin values $\{h_k\}_1^K$ as samples from the continuous function $f(x)$, we see that the rectangular kernel in equation 2 is undesirable, since it implies a violation of the sampling theorem in the expectation sense (see also [1]). On the other hand, an excessively smooth kernel will remove a lot of the details of the distribution $f$. Like in other linear estimation methods we encounter a lower limit on the product of uncertainty in position and in property. For this reason a common trade off is to use a Gaussian kernel, since it is known have the smallest uncertainty product [2].

---

[1]Normally the kernel is symmetric, which makes the mirroring an identity operation.

A kernel with a finite support directly implies a violation of the sampling theorem as mentioned above, but if we choose to look only at peak locations, instead of the entire density function, there is a way around this conflict. The $\cos^2$ kernels used in the *channel representation* [5, 3], both have a compact support and a known inverse with respect to peak locations [1], properties that seem promising for the application to our problem. In the following sections we will make a comparison of peak localisation and interference properties when using $\cos^2$, Gaussian, and the rectangular kernels in equation 2.

# 3 Local Scalar Reconstruction

If we want to detect several peaks in the PDF we have to make a *local scalar reconstruction*, and to minimise the interference between different peaks, the support of the reconstruction should be as small as possible. We will now derive such local scalar reconstructions for both $\cos^2$ functions and Gaussians.

## 3.1 Local $\cos^2$ reconstruction

The channel representation is a unified way to represent a signal $s$, and an associated relevance measure $r$, using localised band-pass functions. Each of the basis functions $\psi_k(s,r)$ has compact support (they are zero along most of the real axis) and raise smoothly to a value $r$ (the relevance measure) as the signal $s$ comes near a specific scalar value $k$:

$$\psi_k(s,r) = \begin{cases} r\cos^2\left(\omega(s-k)\right) & |s-k| < \frac{\pi}{2\omega} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

If we distribute our basis functions with unit distance, i.e. $k \in \mathbb{Z}$, the parameter $\omega$ can be used to control the correlation (or overlap), between neighbouring channel values. For this reason the $\omega$ parameter is called the *channel overlap*.

The local peak can be computed using an idea illustrated in figure 2. The channel values are now seen as samples from an envelope function which peaks at the scalar value $s$. The index of the first active channel is denoted $l$ (in the figure we have $l = 4$). For overlaps of $\omega = \pi/N$ where $N = 3, 4, 5 \ldots$, the number of channels related to a specific scalar becomes[2] $N$ (in the figure we have $N = 3$).

If we assume that the channel values of the $N$ active channels conform to the basis function shape $\psi_k(s,r)$, we obtain $N$ equations:

$$\begin{pmatrix} h_l \\ h_{l+1} \\ \vdots \\ h_{l+N-1} \end{pmatrix} = \begin{pmatrix} \psi_l(s,r) \\ \psi_{l+1}(s,r) \\ \vdots \\ \psi_{l+N-1}(s,r) \end{pmatrix} \tag{6}$$

---

[2]Actually the number of related channels becomes $N-1$ at regularly occurring locations, but this does not affect the resultant reconstruction.
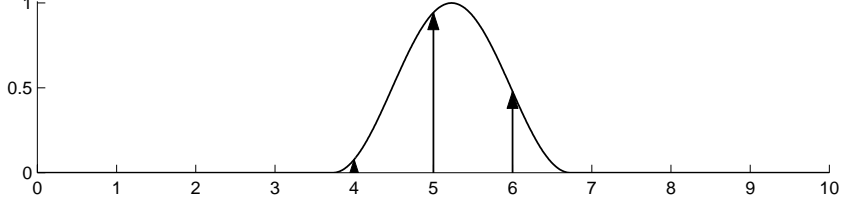
Figure 2: Example of channel values.

*In this example, $\omega = \pi/3$, and $s = 5.23$*

We will now transform an arbitrary row of this system in a number of steps:

$$h_{l+d} = \psi_{l+d}(s, r) = r \cos^2(\omega(s - l - d)) \tag{7}$$

$$h_{l+d} = r/2(1 + \cos(2\omega(s - l - d))) \tag{8}$$

$$h_{l+d} = r/2(1 + \cos(2\omega(s - l))\cos(2\omega d) + \sin(2\omega(s - l))\sin(2\omega d)) \tag{9}$$

$$h_{l+d} = \begin{pmatrix} \frac{1}{2}\cos(2\omega d) & \frac{1}{2}\sin(2\omega d) & \frac{1}{2} \end{pmatrix} \begin{pmatrix} r\cos(2\omega(s - l)) \\ r\sin(2\omega(s - l)) \\ r \end{pmatrix} \tag{10}$$

And thus the entire equation system can be written as:

$$\underbrace{\begin{pmatrix} h_l \\ h_{l+1} \\ \vdots \\ h_{l+N-1} \end{pmatrix}}_{\mathbf{h}} = \frac{1}{2} \underbrace{\begin{pmatrix} \cos(2\omega 0) & \sin(2\omega 0) & 1 \\ \cos(2\omega 1) & \sin(2\omega 1) & 1 \\ \vdots & \vdots & \vdots \\ \cos(2\omega(N-1)) & \sin(2\omega(N-1)) & 1 \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} r\cos(2\omega(s - l)) \\ r\sin(2\omega(s - l)) \\ r \end{pmatrix}}_{\mathbf{d}} \tag{11}$$

This system can be solved using a least-squares fit:

$$\begin{pmatrix} r\cos(2\omega(s - l)) \\ r\sin(2\omega(s - l)) \\ r \end{pmatrix} = \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} = (\mathbf{A}^T\mathbf{A})^{-1}(\mathbf{A}^T\mathbf{h}) \tag{12}$$

Finally, the scalar estimate can be computed as:

$$\hat{s} = l + \frac{1}{2\omega}\arg[d_1 + \boldsymbol{i}d_2] \tag{13}$$

For the relevance estimate, we have two solutions:

$$\hat{r}_1 = |d_1 + \boldsymbol{i}d_2| \quad \text{and} \quad \hat{r}_2 = d_3 \tag{14}$$

4

When they are equal, we have a local interval of of channel values $\{h_l, h_{l+1}, \ldots, h_{l+N-1}\}$ that originate from a single sample $s_n$. Thus, the discrepancy between $\hat{r}_1$ and $\hat{r}_2$ could be used to estimate of the degree of signal model violation.

Now we have to remember that the solution in equation 13 is a *local inverse*. It is thus only valid in a limited range. In [1] (theorem A.1 on page 91) the valid range is shown to be $l + N - 1 - \frac{\pi}{2\omega} \le s \le l + \frac{\pi}{2\omega}$.

For values of $\omega = \pi/N$ where $N = 3, 4, 5 \ldots$, the matrix $\mathbf{A}^T\mathbf{A}$ becomes diagonal, and we can compute the local inverse as a local weighted summation of complex exponentials:

$$\hat{s} = l + \frac{1}{2\omega} \arg \left[ \sum_{k=l}^{l+N-1} h_k e^{\boldsymbol{i}2\omega(k-l)} \right] \tag{15}$$

Although this reconstruction assumes that the channel values $h_k$ have been generated directly from equation 5 it is surprisingly robust to noise, partly due to the fact that scalings will be reflected in the relevance $\hat{r}$ alone.

The assumption that the channel values originate from $\cos^2$ functions with a certain $\omega$ value is unfortunate, since the shape of the reconstructed function should look like the PDF convolved with the kernel function (see section 2). If we have high amounts of measurement noise, the PDF peak will become less distinct, and the channels will tend to have a larger de facto overlap than what is indicated by the kernel function, leading to systematic errors in the peak estimation.

One way to deal with this problem, at least partially could be to perform the reconstruction for a set of different $\omega$ values, and choose the reconstruction with the smallest difference $\hat{r}_2 - \hat{r}_1$.

## 3.2 Local Gaussian reconstruction

If we use Gaussian kernels, the contribution to a bin from a single sample looks like this:

$$\varphi_k(s,r) = h_k = re^{-\dfrac{(s-k)^2}{2\sigma^2}} \tag{16}$$

Here we have added a parameter $r$ that corresponds to the sample relevance, i.e. we allow each sample to be assigned a weight. This is normally not done in PDF estimation, but we introduce it here to show the similarities to the channel representation in the previous section. Combined with a local reconstruction, the introduction of a relevance measure is all that is needed for Gaussians to become an alternative to $\cos^2$ functions in associative learning.

If we look at three neighbouring bins around $k = l$, we obtain three equations:

$$\begin{pmatrix} h_{l-1} \\ h_l \\ h_{l+1} \end{pmatrix} = \begin{pmatrix} \varphi_{l-1}(s,r) \\ \varphi_l(s,r) \\ \varphi_{l+1}(s,r) \end{pmatrix} \tag{17}$$

5

The logarithm of an arbitrary row can be written as:

$$\ln h_{l+d} = \ln r - \frac{(s-l-d)^2}{2\sigma^2} \tag{18}$$

$$= \begin{pmatrix} 1 & d & d^2 \end{pmatrix} \underbrace{\begin{pmatrix} \ln r - \frac{(s-l)^2}{2\sigma^2} & \frac{s-l}{\sigma^2} & -\frac{1}{2\sigma^2} \end{pmatrix}^T}_{\mathbf{p}} \tag{19}$$

We now have an equation system of the form:

$$\ln \mathbf{h} = \mathbf{D}\mathbf{p} \tag{20}$$

with the solution

$$\mathbf{p} = \underbrace{\begin{pmatrix} 0 & 1 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & -1 & \frac{1}{2} \end{pmatrix}}_{\mathbf{D}^{-1}} \ln \mathbf{h} \tag{21}$$

From the solution $\mathbf{p}$ we can find the estimates $\hat{s}$, $\hat{\sigma}$, and $\hat{r}$ as:

$$\hat{s} = l - \frac{p_2}{2p_3} \qquad \hat{\sigma} = \sqrt{-\frac{1}{2p_3}} \qquad \hat{r} = e^{p_1 - \frac{p_2^2}{4p_3}}$$

It is interesting to note that we can now find a direct estimate of standard deviation for the measurement noise (using the addition theorem for variances) as $\hat{\sigma}_{\text{noise}} = \sqrt{\hat{\sigma}^2 - \sigma^2}$. The $\hat{r}$-value of the reconstruction can be seen as a measure of how much sample relevance we have near the reconstructed peak. For a high confidence in a peak location we should thus have a small $\hat{\sigma}_{\text{noise}}$ and a large relevance $\hat{r}$.

Note that the $\sigma$-parameter is estimated directly here, contrary to the $\omega$ parameter of the $\cos^2$ channels, which is assumed to be known a priori. To be able to estimate the dispersion directly is a potential advantage.

# 4    Noise Generation

A uniform noise can be generated from the following distribution:

$$f_{\mathcal{R}}(x) = \begin{cases} 1/a & \text{when} \quad -a/2 \leq x \leq a/2 \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

Where $a$ is a parameter that controls the variance of the distribution. Unit variance is obtained when $a = \sqrt{12}$, and thus an arbitrary standard deviation $\sigma$ can be obtained by setting $a = \sigma\sqrt{12}$.
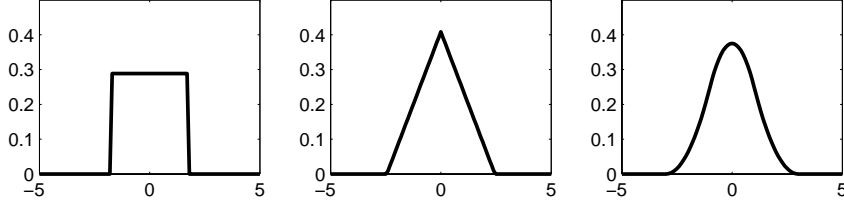
Figure 3: PDFs of $\mathcal{D}^1$, $\mathcal{D}^2$, and $\mathcal{D}^3$ with $\sigma = 1$.

We can generate increasingly Gaussian-like noise as sums of samples drawn from $\mathcal{R}$:

$$d^K = \sum_{k=1}^{K} r_k \quad \text{where} \quad r_k \in \mathcal{R} \tag{23}$$

The obtained distribution $\mathcal{D}^K$ has a PDF that equals the convolution of the individual PDFs. For instance we have $f_{\mathcal{D}^3}(x) = (f_{\mathcal{R}} * f_{\mathcal{R}} * f_{\mathcal{R}})(x)$. Figure 3 shows the obtained distributions $\mathcal{D}^1$, $\mathcal{D}^2$, and $\mathcal{D}^3$. As can be seen, we are able to successively move closer to a normal distribution by increasing $K$. In fact, according to the central limit theorem we obtain a normal distribution at the limit $K \to \infty$ (see e.g.[4]).

The variance of a sum equals the sum of the individual variances. This means that

$$\text{var}(\mathcal{D}^K) = K \times \text{var}(\mathcal{R}) \tag{24}$$

and thus we can control the standard deviation of $\mathcal{D}^K$ by setting $a = \sigma\sqrt{12/K}$.

To illustrate the histogram generation we have generated a set of $N = 1000$ samples from $\mathcal{D}^1$ and computed histograms of $s = d_1 + l$ for $l = 4, 4.25, 4.5, 4.75$. The result is shown in figure 4.
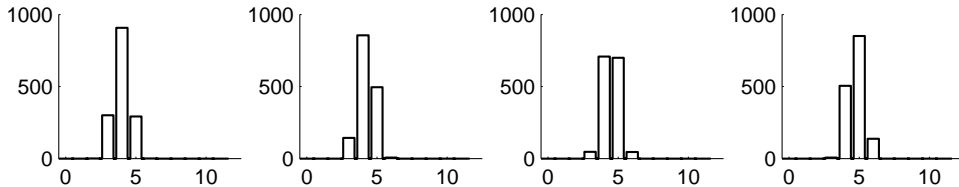


Figure 4: Histograms from $\cos^2$ kernels.
*Left to right: $\cos^2$ histograms with $\omega = \pi/3$ and samples centred around 4, 4.25, 4.5, and 4.75. Noise is $\mathcal{D}^1$ with $\sigma = 0.3$. The histograms use bins with centres at 0, 1, ...12.*

The estimated peaks using equation 15 are: 4.0068, 4.2497, 4.5073, and 4.7436. As a definition of the correct peak location we will use the sample average:

$$\bar{s} = \frac{1}{N} \sum_{n=1}^{N} s_n \tag{25}$$

The corresponding sample averages are 4.0059, 4.2474, 4.5079, and 4.7467. As we can see, the peak detection is quite accurate.

# 5 Experiments on varied noise and bin shape

In the first experiment we will look at how the number of samples influence the accuracy of the peak detection. We set each sample to be

$$s_n = l + d_n \quad \text{where} \quad d_n \in \mathcal{D}^3 \tag{26}$$

The offset $l$ is varied in the interval $[4, 6]$ with steps of 0.005, giving a total of 401 different positions. We then vary the number of samples $N$ between 1 and 100, and compute the root-mean-squared-error (RMSE) between the estimated peak location and the sample mean. The results are shown in figure 5 for $\cos^2$ kernels (left) and Gaussian kernels (right). Thin dotted curves show the error for rectangular bins as a comparison.
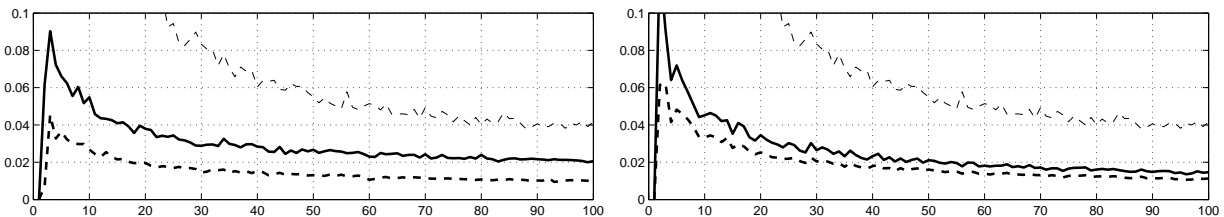


Figure 5: RMSE under varied number of samples.
*Left:* $\cos^2$ *histograms with* $\omega = \pi/3$ *(solid) and* $\omega = \pi/4$ *(dashed).* *Right: Gaussian histograms with* $\sigma = 0.60$ *(solid) and* $0.80$ *(dashed).* *Noise is* $\mathcal{D}^3$ *with* $\sigma = 0.5$. *Each error is the average over* 401 *positions. Thin dotted curves show the error for rectangular bins as a comparison.*

As these plots hopefully illustrate, the error levels off at a value above zero. This indicates that we have an inherent error in the peak detection algorithm. As can be expected, the error for the rectangular histograms is considerably larger than for the other methods, and as we shall see later, the chosen noise level of $\sigma = 0.5$ actually reduces the error for this bin shape drastically. The Gaussian kernels used in the right part of figure 5 have $\sigma$ values 0.6 and 0.8. These are chosen to give kernel functions with the same area as the $\cos^2$ kernels with $\omega$ values $\pi/3$ and $\pi/4$ tried in the left plot. Same areas for the kernels are obtained when $\sigma = \sqrt{\pi/8}/\omega$.

## 5.1 Position Dependency of Estimation Errors

We will now look at how the peak estimation error varies with peak position. Just like in the previous experiment we let $l$ vary in the interval $[4, 6]$ with steps of 0.005. The estimation error $\hat{s} - \bar{s}$ is plotted against the sample averages $\bar{s}$ in figure 6. As can be seen, there is a periodic variation with position. Apparently the error is smaller for peaks near integer positions, which is where the bin centres are located.

By increasing the number of samples, we can get rid of the random disturbance, and see the shape of the periodic error function better. This is done in figure 7. We have tried the three noise distributions $\mathcal{D}^1$, $\mathcal{D}^2$, and $\mathcal{D}^3$ with $\sigma = 0.3$. The error for the $\cos^2$

8

and a Gaussian with corresponding area are compared. For the $\cos^2$ we can see that the shape of the error function changes with the noise, with a slight preference for the uniform distribution $\mathcal{D}^1$. The maximum error however, appears to stay the same. For the Gaussian on the other hand, we can see that the error drops sharply as the error tends toward a normal distribution.
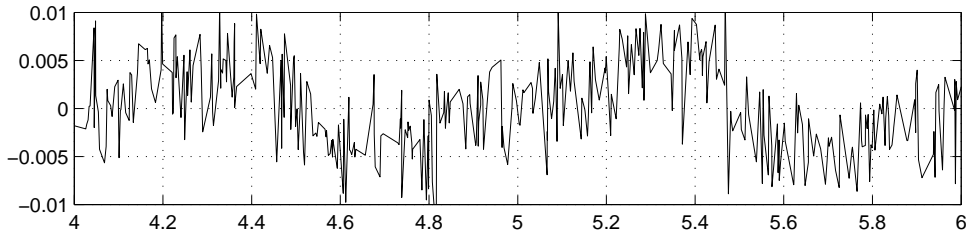


Figure 6: Position dependency of error in peak estimation.
*Error in peak detection for $\cos^2$ histograms with $\omega = \pi/3$. Noise is $\mathcal{D}^3$ with $\sigma = 0.3$. Position is varied in steps of $0.005$. Number of samples $N = 100$.*
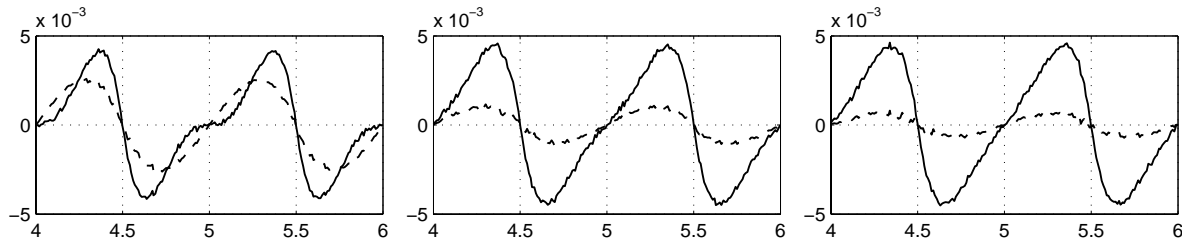


Figure 7: Position dependency of error in peak estimation.
*Left to right: Noise $\mathcal{D}^1$, $\mathcal{D}^2$, and $\mathcal{D}^3$ with $\sigma = 0.3$. Solid curves are $\cos^2$ bins with $\omega = \pi/3$. Dashed curves are Gaussian bins with $\sigma = 0.6$. Position is varied in steps of $0.01$. Number of samples $N = 100\,000$.*
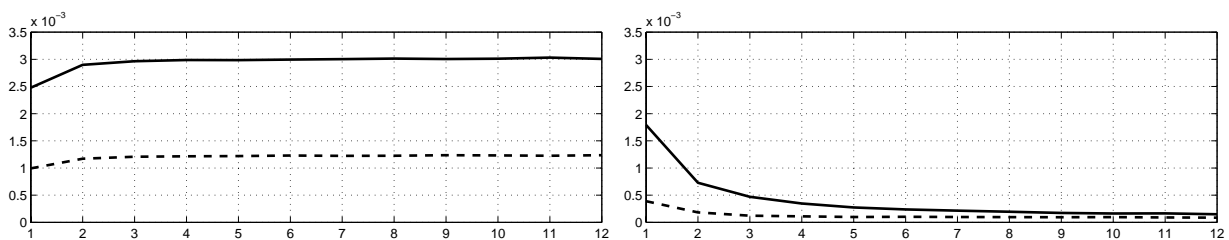


Figure 8: Noise shape dependency of error in peak estimation.
*Left: $\cos^2$ bins with $\omega = \pi/3$ (solid) and $\omega = \pi/4$ (dashed). Right: Gaussian bins with $\sigma = 0.6$ (solid) and $\sigma = 0.8$ (dashed). Noise is varied from $\mathcal{D}^1$, through $\mathcal{D}^{12}$ with $\sigma = 0.3$. Position is varied in steps of $0.01$. Number of samples $N = 100\,000$.*

These trends continue as the noise tends toward the normal distribution. Figure 8 shows the RMS error for $\mathcal{D}^1$ through $\mathcal{D}^{12}$ with $\sigma = 0.3$. Note however that the very low

estimation errors for Gaussian bins with near normal distributed noise are due to the very large number of samples ($N = 100\,000$) used in this experiment. For lower number of samples the curves level off at higher error rates.

## 5.2   Varied noise level

We will now have a look at what happens when the noise level is varied. Like before, the offset is varied in the range $[4, 6]$ in steps of $0.005$. The noise distribution is set to $\mathcal{D}^3$, with the standard deviation varied in the range $[0, 1]$ in steps of $0.01$.

The left and centre plots of figure 9 shows the error for $\cos^2$ bins and Gaussian bins with the corresponding reconstruction. As can be expected, the peak accuracy decreases with the noise level. We can also see that larger kernel support appears to give slightly smaller errors. However, most interesting is perhaps the behaviour for the rectangular bins shown in the right plot. As can be seen in the plot, the optimal noise level is actually above zero. This is due to the expectation of the bin values being the convolution of the kernel and the noise PDF (see section 2). The added noise thus results in a smoother PDF sampling. The same effect motivates the use of dithering noise for relaxation of quantisation errors [1]. Finding the optimal choice of noise given a kernel function will be called *the dithering problem.*
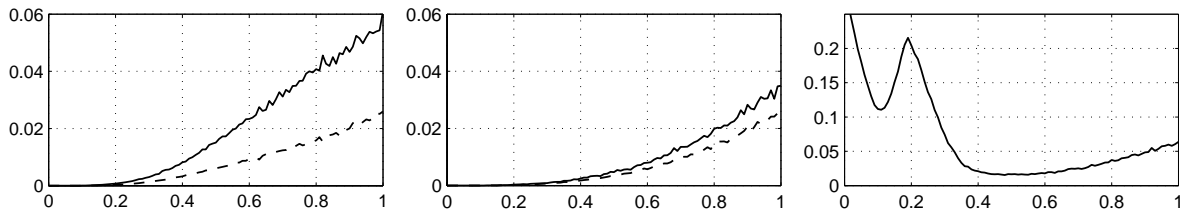


Figure 9: RMSE of reconstruction as function of noise level.
*Left: $\cos^2$ bins with $\omega = \pi/3$ (solid) and $\omega = \pi/4$ (dashed). Centre: Gaussian bins with $\sigma = 0.6$ (solid) and $\sigma = 0.8$ (dashed). Right: rectangular bins (note the different scaling of the y-axis). Noise is $\mathcal{D}^3$, with $\sigma \in [0, 1]$. Position is varied in steps of $0.01$. Number of samples $N = 1\,000$.*

Another interesting observation from figure 9 is that above a certain noise level, the three methods have errors of the same order of magnitude. For a given non-zero level of sensor noise, there appears to be considerable freedom in choosing kernel shape and still get small errors. Since a small support for a kernel is also desirable, due to reduced interference between competing hypotheses, one could actually suspect that the optimal kernel among these is the one with the smallest support. Such a kernel would of course cause considerable aliasing in the noise-free case. The choice of kernel given a certain noise can be seen as the dual to the dithering problem, and will thus be called *the inverse dithering problem.*

Biological neurons are known have binary responses (i.e. at a given time instant they either fire or don't fire). They are able to convey graded information by having the rate of firing depend on the sum of the incoming (afferent) signals. This behaviour could be modelled as (temporally local) histogram computations with noise added before

application of the bin function. If the temporal averaging in the neurons is larger than just a few samples, it would be reasonable to expect that biological neurons implicitly have solved the inverse dithering problem.

## 5.3  Aliasing due to small support

For small bin sizes the local Gaussian reconstruction will encounter an aliasing-like effect. To demonstrate this we will now encode single scalars and reconstruct them. We vary the bin size parameter $\sigma$ in the range $[0, 0.5]$ in steps of 0.002. For each $\sigma$-value, the spatial position is varied in the range $[5, 6]$ in steps of 0.05. The absolute error of the reconstruction is averaged over spatial position, and plotted against the $\sigma$-value in the left part of figure 10. The position dependency of the error is shown in the right part of the figure.
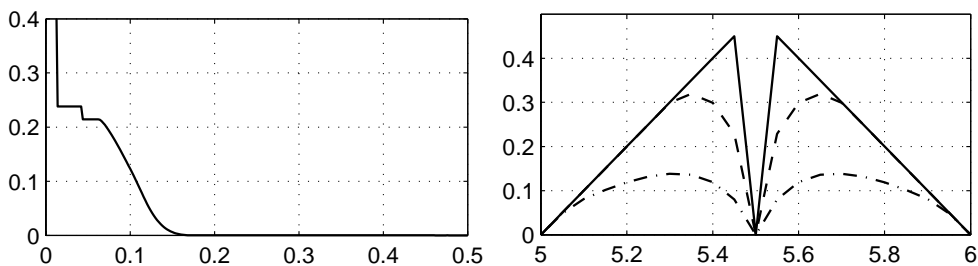


Figure 10: Aliasing for small bin sizes.
*Left: Average absolute reconstruction error as function of bin size. Right: absolute reconstruction error as function of scalar position. $\sigma = 0.05$ (solid), $\sigma = 0.08$ (dashed) $\sigma = 0.11$ (dash-dot).*

As can be seen from this experiment, the reconstruction error is negligible for bin sizes larger than $\sigma = 0.2$. For smaller bin sizes, the reconstruction becomes increasingly more unfair toward positions far from integer or half-integer positions. Note that this experiment also corresponds to the case of Gaussian bins and Gaussian noise in the expectation sense (see section 2).

# 6  Experiments on peak interference

The purpose of the local estimation is detection of multiple peaks, and we will now have a look at how the presence of other peaks influence the estimation.

We construct a new variable $s_n$ as:

$$s_n = l_1 b_n + l_2(1 - b_n) + d_n \quad d_n \in \mathcal{D}^3 \quad b_n \in \{0, 1\} \tag{27}$$

where $b_n$ is a binary variable, assuming the values 0 and 1 with equal probability, and $d_n \in \mathcal{D}^3$. The variable $s_n$ should have marked density peaks near the values $l_1$ and $l_2$.

Initially we will have a constant position $l_1 = 5.5$ and vary the position of $l_2$ in the range $[1, 10]$ with steps of 0.1. As an illustration of the peak interference we will now compute

11

histograms from $\cos^2$ kernels ($\omega = \pi/3$), with bin centres $c_k = \{-1, 0, \dots 11, 12\}$, and plot the estimated peak positions (see figure 11, right).
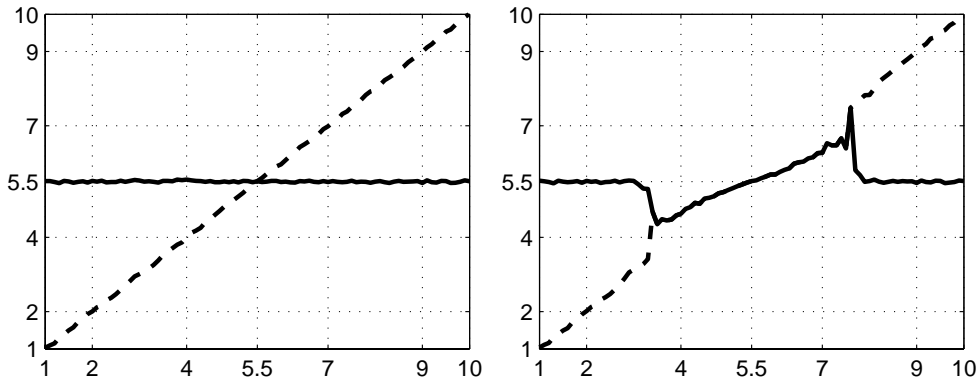


Figure 11: Interference for $\cos^2$ kernels.

*Left: $\bar{s}_1$ (solid) $\bar{s}_2$ (dashed) plotted against $l_2$. Right: estimated peak locations, peak closest to $l_1$ (solid) and closest to $l_2$ (dashed), plotted against $l_2$. Bins are $\cos^2$ with $\omega = \pi/3$, noise is $\mathcal{D}^3$ with $\sigma = 0.5$, and number of samples $N = 1\,000$*

In order to estimate the accuracy of the peak detection we will also compute the reference averages $\bar{s}_1$ and $\bar{s}_2$ as the averages of $s_n$ when $b_n$ is equal to 0 and 1 respectively:

$$\bar{s}_1 = \frac{\sum (l_1 + d_n) b_n}{\sum b_n} \qquad \bar{s}_2 = \frac{\sum (l_2 + d_n)(1 - b_n)}{N - \sum b_n} \tag{28}$$

The sample averages are plotted in figure 11, left. As we can see from the figure, the peaks are not distinguished if $|l_1 - l_2| < 2$. This minimum distance is called the *metameric distance* in [1]. In the experiments to follow, we will investigate the metameric distance in more detail.

## 6.1  Space variant metameric distance

We are now going to investigate how the metameric distance behaves under varying position of the peaks. Due to the symmetry of the generated signal (see equation 27), $\{l_1 = a, l_2 = b\}$ is equivalent to $\{l_1 = b, l_2 = a\}$. Thus it will suffice to let $l_1$ vary in the range $[1, 10]$, and $l_2$ in $[l_1, 10]$. The missing results $\{l_1 = a, l_2 = b, b < a\}$ can be copied from the equivalent $\{l_1 = b, l_2 = a\}$ if needed.

We start by redoing the experiment of section 6, but this time we vary both $l_1$ and $l_2$. The absolute errors for the reconstructed scalars are shown in figure 12, left and centre. The sum of the errors is shown in the right plot of the figure. As can be seen from this experiment, the metameric distance is space variant. The metameric distance is the largest when one of the peaks is located exactly between two bin centres.
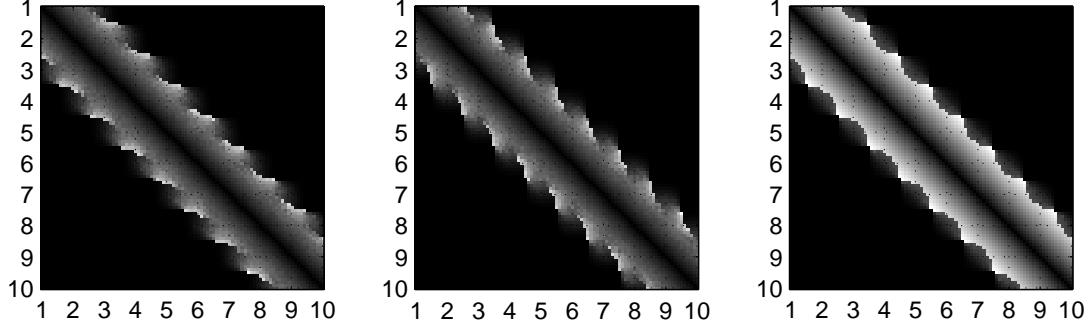
12

Figure 12: Absolute error for $\cos^2$ kernel with varying peak positions.
*Left to right: Absolute errors for reconstructions closest to $l_1$, and $l_2$, and the error sum. High intensities correspond to large errors. Bins are $\cos^2$ with $\omega = \pi/3$, noise is $\mathcal{D}^3$ with $\sigma = 0.5$, and number of samples $N = 1\,000$*

## 6.2 Varied bin overlap

We will now demonstrate how the metameric distance varies with the overlap of the bins. In order to isolate the effect of the bin overlap, we will only encode two scalars, sum them, and reconstruct. Note that this also corresponds to the expectation values of Gaussian bins with Gaussian noise according to section 2. We vary the locations $l_1$ and $l_2$ of the scalars in the range $[3, 8]$ in steps of 0.02, and vary the bin size $\sigma$ in the range $[0.1, 1.2]$ in steps of 0.1. The results are shown in figure 13.
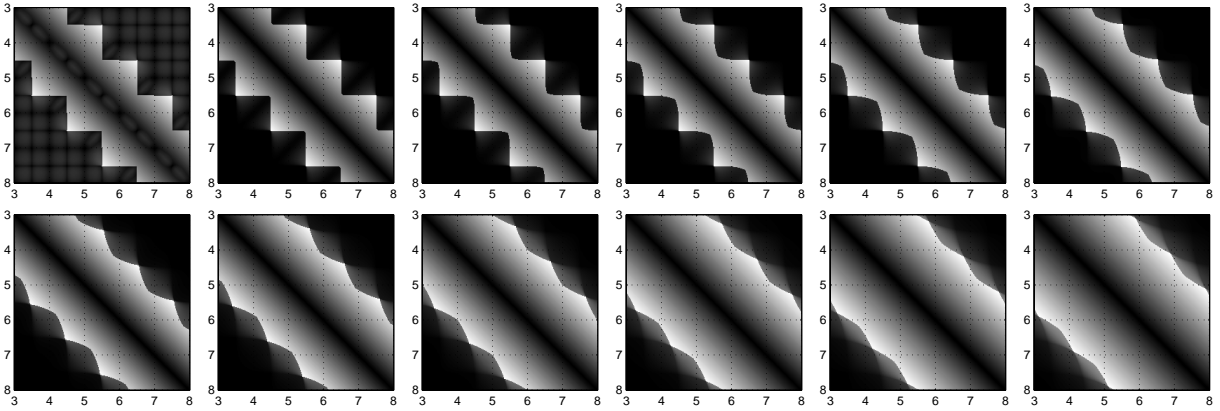


Figure 13: Reconstruction errors for Gaussian kernels.
*Left to right, top to bottom: Sum of absolute reconstruction errors for Gaussian kernels with $\sigma = 0.1, 0.2 \ldots 1.2$*

As can be seen from these plots, the metameric distance becomes less space variant as we increase the bin overlap. We can also note that the largest metameric distance is no longer found when one of the peaks is between two bin centres for bin sizes above $\sigma = 0.5$. A good choice might thus be to stop at $\sigma = 0.5$. This value also gives $\frac{\partial^2}{\partial s^2}(\varphi_k(s) + \varphi_{k+1}(s)) = 0$ for $s = k+0.5$), which means that the sum of the bin values is less dependent on the actual location of the peak. We can also note that the reconstruction aliasing

investigated in section 5.3 is visible in the first plot.

## 6.3   Other bin shapes

As a comparison to the Gaussian kernels we now try the same experiment with $\cos^2$ ($\omega = \pi/3$, $\pi/4$, $\pi/5$, and $\pi/6$), and rectangular kernels. The results are shown in figure 14.
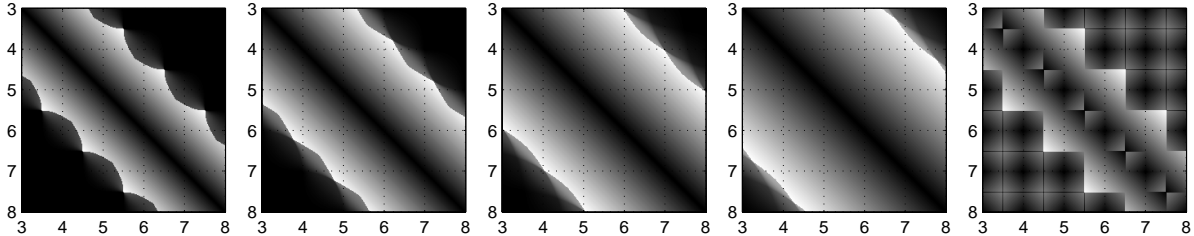


Figure 14: Reconstruction errors for $\cos^2$ and rectangular kernels.
*Left to right: Sum of absolute reconstruction errors for* $\cos^2$ *($\omega = \pi/3$, $\pi/4$, $\pi/5$, and $\pi/6$) and rectangular kernels.*

If we compare the reconstruction errors in figure 14 with those in figure 13, we can note that the behaviour for $\omega = \pi/3$ is roughly equivalent to $\sigma = 0.6$, while the metameric distance is significantly larger for $\omega = \pi/4$, $\pi/5$, and $\pi/6$ than for the corresponding $\sigma$-values (0.8, 1.0, and 1.2). This is explained by the fact that the reconstruction for $\cos^2$ bins use increasingly more adjacent values in the reconstruction, while the Gaussian reconstruction always uses three.

As can be seen from the rightmost plot in figure 14, the metameric distance for the rectangular bins is quite small. However, the overall localisation ability is considerably affected by this bin shape. The rectangular bins can thus be said to allow peaks to be told apart at smaller distances at the price of an overall less exact localisation. Note however that the aliasing effect is avoided for the Gaussian kernels in figure 13, while still keeping the metameric distance down, so this is not a motivation for the use of rectangular bins.

# 7   Concluding remarks

A HiperLearn network [3] is a linear mapping using signals in *the channel representation* (see section 3.1). The response signal model for a HiperLearn network is more complicated than the sum of a discrete and a continuous localised distribution used in this report. However, the networks use localised response functions, which can roughly be seen as corresponding to the localised kernels used in this report. The local reconstruction presented here is identical to the one used in HiperLearn networks in the $\cos^2$ kernel case, and the Gaussian kernels presented here is the missing component that allows the Gaussian kernels to be used as an alternative to $\cos^2$ functions in HiperLearn.

It is also interesting to note that the optimisation of the linkage matrix in HiperLearn is related to a discrete version of the inverse dithering problem (see section 5.2). To be

more precise, we have to view each weighted sum of feature function values as a sample from the distribution, and since the response functions are prescribed in HiperLearn we end up with an equivalence to a deconvolution in the expectation sense.

## Acknowledgements

## References

[1] Per-Erik Forssén. Sparse Representations for Medium Level Vision. Lic. Thesis LiU-Tek-Lic-2001:06, Dept. EE, Linköping University, SE-581 83 Linköping, Sweden, February 2001. Thesis No. 869, ISBN 91-7219-951-2.

[2] D. Gabor. Theory of communication. *J. Inst. Elec. Eng.*, 93(26):429–457, 1946.

[3] Gösta Granlund, Per-Erik Forssén, and Björn Johansson. HiperLearn: A High Performance Learning Architecture. Technical Report LiTH-ISY-R-2409, Dept. EE, Linköping University, SE-581 83 Linköping, Sweden, January 2002.

[4] Aapo Hyvärinen, Juha Karhunen, and Erkki Oja. *Independent Component Analysis*. John Wiley and Sons Inc., 2001. ISBN 0-471-40540-X.

[5] K. Nordberg, G. Granlund, and H. Knutsson. Representation and Learning of Invariance. In *Proceedings of IEEE International Conference on Image Processing*, Austin, Texas, November 1994. IEEE.