# Blind restoration and superresolution using generalized cross-validation with Gauss quadrature rules

Nhat Nguyen*, Peyman Milanfar†, and Gene Golub‡

February 11, 2000

EDICS numbers: 2-SEQP, 2-MLFT, 2-ANAL

**Abstract**

In many image restoration/superresolution applications, the blurring process, i.e., point spread function (PSF) of the imaging system, is not known or known only to within a set of parameters. We estimate these PSF parameters for this ill-posed class of inverse problem from raw data, along with the regularization parameters required to stabilize the solution, using the generalized cross-validation method (GCV). To reduce the computational complexity of GCV, we propose efficient approximation techniques based on the Lanczos algorithm and Gauss quadrature theory. Data-driven blind restoration/superresolution experiments with synthetic and Forward Looking Infrared (FLIR) image sequences are presented to demonstrate the effectiveness and robustness of our method.

*Nhat Nguyen and Gene Golub are with Scientific Computing and Computational Mathematics Program, Stanford University, Stanford CA 94305-9025, {*nguyen,golub*}*@sccm.stanford.edu*

†Peyman Milanfar is with the Deparment of Electrical Engineering, University of California, Santa Cruz, CA 95064, *milanfar@cse.ucsc.edu*

‡The work in this report was supported in part by the National Science Foundation Grant CCR-9505393.

# 1 Introduction

Image superresolution refers to image processing algorithms which produce high quality, high-resolution (HR) images from a set of low quality, low-resolution (LR) images. While there is always a demand for better quality images, the level of image detail is crucial in the performance of several computer vision algorithms. Target recognition, detection and identification systems are some of the military applications which require highest quality achievable images. License plate readers, surveillance monitors, and medical imaging applications are examples of civilian applications with the same requirement. In many visual applications, both civilian and military, the imaging sensors have poor resolution outputs. When resolution can not be improved by replacing sensors, either because of cost or hardware physical limits, we can resort to superresolution algorithms. Even when superior equipment is available, superresolution algorithms can provide an inexpensive alternative.

Superresolution, at its core, is a process by which one gains spatial resolution in return for temporal bandwidth. Temporal bandwidth refers to availability of multiple images of the same scene. Lukosz [21, 22] had realized the possibity of superresolution by trading off temporal for spatial resolution. However, superresolution can not perform miracles. We can not expect to extract subpixel information from a sequence of identical images as there is effectively no temporal bandwidth to trade-off. There must be nonredundant information among the images. We must be able to translate data temporal bandwidth into subpixel image content. Each LR frame provides a different "look" at the same scene. Theoretically, by providing different lighting conditions or with different sensors, superresolution can be achieved without relative scene motion. This is the multichannel data fusion superresolution problem [31]. With just one imaging device and under the same lighting conditions, we require that there be some relative motion from frame to frame. Frame to

frame motion can be a combination of camera platform motion relative to the scene, moving objects in the scene, and camera jitters. For example, in satellite imaging, images of the ground below are captured as the camera orbits the earth, whereas in surveillance and monitoring applications, the camera is placed on a fixed platform, and observed objects move within the scene. Motion and nonredundant information are what make superresolution possible. Armed with this information, we are able to extract subpixel content at a higher resolution than in each individual frame.

The rest of the paper is organized as follows. In Section 2 we outline the problem description and model for superresolution. Image restoration is a special case of superresolution, and Section 3 shows the relationship between them. Techniques for blind superresolution proposed in this paper can be applied equally well to blind restoration. Section 4 describes the blur identification problem and our approach using the generalized cross-validation (GCV) method in detail. In Section 6, we present our method for bounding the GCV criterion value accurately and efficiently, based on quadrature rules and the Lanczos algorithm. Blur estimation and blind superresolution results are shown in Section 7.

## 2 Problem Description

In this section, we describe a straightforward and efficient model for superresolution which will serve as the foundation for the development of algorithms for the rest of the paper. Although superresolution and multichannel data fusion can both be described with one model, for simplicity and ease of presentation, our superresolution representation will consist of LR data from one imaging source. We assume consistent lighting conditions, negligible optical distortions, that objects observed are acquired under orthographic projections, and that individual scene motions can be modeled, or approximated, as affine transformations. While these simplifications

result in some loss of generality, the model will be adequate for the majority of superresolution imaging applications.

The problem can be stated as follows:

> *Given a set of degraded LR frames* $\{\mathbf{f}_k\}_{k=1,\ldots,p}$ *each* $M \times N$ *pixels in dimension under the conditions above, and a desired enhancement factor* $r$*, reconstruct an enhanced/restored HR image with dimensions* $rM \times rN$*.*

Figure 1 illustrates the problem setup. The figure shows three $4 \times 4$ pixel LR frames on an $8 \times 8$ HR grid. Each symbol (square, circle, triangle) indicates the sampling points of a frame with respect to the HR grid. We pick an arbitrary frame as a reference frame; in this case, the frame marked by the circular symbols. The sampling grid for the triangular frame is a simple translation of the reference frame grid. The difference between the sampling grid for the square frame and the reference frame grid includes translational, rotational, and magnification (zoom) components. The goal of superresolution is to interpolate and restore values at the HR grid points.

In order to solve for the unknown HR values, we first model the forward process, which takes the ideal HR image to a degraded LR frame. Many imaging devices today utilize charged-coupled devices (CCD) which are arrays of light detectors. A detector determines pixel intensity values depending upon the amount of light detected from its respective area in the scene. Resolution of images produced by the camera is proportional to the density of the detector array. Figure 2 shows a simplistic model of a CCD camera.

The camera lens and aperture produce a blurred version of the object. The CCD array turns this degraded analog signal into a discrete 2-D image. In addition, the images are contaminated by additive noise from various sources: quantization errors, sensor measurement, model errors, etc. Ideally, we would like to produce an

HR image corresponding to placing a high density CCD array placed *in front of* the camera lens. The forward relationship between a degraded, LR frame and the ideal HR image can be described as follows [8]:

$$\begin{aligned} \mathbf{f}_k &= DC_kE_k\mathbf{x} + \mathbf{n}_k, \quad 1 \leq k \leq p \qquad (1) \\ &= H_k\mathbf{x} + \mathbf{n}_k, \qquad (2) \end{aligned}$$

where $D$ is the down-sampling operator, $C_k$'s are the blurring/averaging operators, $E_k$'s are the affine transforms which map the HR grid coordinate system to the LR grid systems, $\mathbf{x}$ is the unknown ideal HR image, and $\mathbf{n}_k$'s are the additive noise vectors. The LR frames $\mathbf{f}_k$ are given, and the decimation operator $D$ is known. The blurring operator $C_k$ and camera lens characteristics are in general unknown. However, the blurring process can be well approximated as linear spatially invariant (LSI). The scene motions for each frame relative to the reference frame are also generally not known. Finally, with multiple independent sources of error, the central limit theorem allows us to assume white Gaussian distribution for the additive noise vectors $\mathbf{n}_k$ with possibly unknown variance. Each frame $\mathbf{f}_k$, $M \times N$ pixels in dimension, becomes a column $MN \times 1$ vector by columnwise reordering. Pixel $(1,1)$ of the 2-D frame is ordered first, pixel $(1,2)$ is second, and so forth. The unknown ideal image $\mathbf{x}$ is reshaped into a $r^2MN \times 1$ column by the same columnwise ordering. The matrices $E_k$ are square $r^2MN \times r^2MN$ matrices representing the affine transforms applied to the ideal image. The matrices $C_k$, also $r^2MN \times r^2MN$ square matrices, represent the blurring operators, and $D$ is an $MN \times r^2MN$ decimation matrix. By stacking the frame equations (2) we get

$$\mathbf{f} = H\mathbf{x} + \mathbf{n}, \qquad (3)$$

where $\mathbf{f}, \mathbf{n}$ are now $pMN \times 1$ vectors and $H$ is the complete system matrix with dimensions $pMN \times r^2MN$. The shape of the system depends on the number of

available LR frames. If $p < r^2$, we have an underdetermined system. If $p = r^2$, the system will be square. And if $p > r^2$, we have an overdetermined system. Techniques developed in this paper are applicable for all three cases. The matrix $H$ is structured, typically ill-conditioned, and very large. The dimensions of $H$ are directly related to the number of data samples and unknowns, which are usually in the tens of thousands. Solving the nonblind superresolution problem (3) is a formidable computing challenge, which we have addressed in [24, 25, 27].

In order to extract subpixel information content from an image sequence, each frame must be registered accurately with respect to some reference to within subpixel accuracy. In this paper, we assume that motion estimation has been performed so that image registration parameters are known explicitly. For a comprehensive review on the topic, the reader is referred to the survey by Brown [6].

## 3 Restoration and Superresolution

Image restoration and image superresolution are closely related problems. In image restoration, the goal is to reconstruct the original image given a degraded image *at the same resolution scale*. Image superresolution reconstructs a higher resolution, restored image from several aliased, degraded, *low resolution* frames. General restoration problems can be modeled as

$$\mathbf{f} = C\mathbf{x} + \mathbf{n}, \tag{4}$$

where $\mathbf{f}$ is the observed degraded image, $\mathbf{x}$ is the original image we wish to estimate, $\mathbf{n}$ consists of various additive system noise sources, and $C$ is the convolution operator representing the blur. In most instances, blurring is assumed to be linear shift invariant. We recover the original image $\mathbf{x}$ by deconvolving the blur $C$ from the degraded observed $\mathbf{f}$. This classical problem has been thoroughly studied and can be

solved by several well-known techniques such as Wiener filtering, recursive Kalman filtering, and iterative deconvolution methods (cf. [1, 34, 4]).

Superresolution includes restoration as a special case. Namely, equation (4) can be rewritten as

$$\mathbf{f}_k = DCE_k\mathbf{x} + \mathbf{n}_k, \quad 1 \leq k \leq r^2, \tag{5}$$

where $r$ is an arbitrary decimation factor, equivalent to the resolution enhancement factor, the LR "frames" $\mathbf{f}_k$'s are the degraded data image having been shifted horizontally and vertically by multiples of one HR pixel and down-sampled by a factor of $r$, and $E_k$'s represent the relative motion shifts. Thus, image restoration can be restated as an image superresolution problem with a desired enhancement factor of $r$ and a set of $r^2$ LR frames with all possible horizontal and vertical HR pixel shifts to cover the entire HR grid. Figure 3 illustrates this process, where in this example, $r = 2$. The noisy, blurred image is partitioned into four LR "frames" each marked by a distinct symbol. Using the frame marked by the circle symbol as reference, the frame marked by the square symbol contains sampling points at one HR pixel shifted in the horizontal direction. Similarly, the triangles mark sampling points at one HR pixel shifted in the vertical direction, and the diamonds sampling points at one HR pixel shifted diagonally. Image restoration is therefore simply image superresolution with regularly sampled LR data completely covering the HR grid. As a result, techniques developed here for blind superresolution will be applicable for blind restoration problems as well.

## 4 Blur Identification

In many practical applications, the blurring process is not known or known only to within a set of parameters. The problem of restoring the original image from

a degraded observation and incomplete information about the blur is called blind deconvolution. Analogously, blind superresolution is superresolving from LR data with incomplete information about the degradation process. There has been extensive work on blind deconvolution. A good survey on the topic can be found in the paper by Kundur and Hatzinakos [19]. Existing blind deconvolution methods can be categorized into two main classes: methods which separate blur identification as a disjoint procedure from restoration, and methods which combine blur identification and restoration in one procedure. Methods in the first class tend to be computationally simpler. Blind deconvolution methods can be generalized to handle multiple observations. Multiframe blind deconvolution is better at suppressing noise and edge artifacts and preventing PSF estimates from converging to the trivial delta function.

Technically, blind deconvolution is a factorization problem of two-dimensional polynomials in the Z-transform domain (see e.g. [19]). Based on this observation and the fact that multidimensional polynomials are generally not factorizable, Lane and Bates [20] find the blurring and original image polynomial factors by examining the roots of the polynomial of the observed image in the Z-transform domain. Although conceptually attractive, zero sheet separation is highly sensitive to noise. Using multiple LR frames, Shekarforoush and Chellappa [31] proposed a related approach of estimating the optical transfer function (OTF) by finding spikes in the magnitude of the cross power spectrum of consecutive frames. Several researchers have considered iterative blind restoration. The most popular of these is the iterative blind deconvolution (IBD) method by Ayers and Dainty [2]. IBD simultaneously reconstructs the blur and image values by alternately enforcing constraints in the image and Fourier domain until estimates for both converge. Biggs and Andrews [5] extended the IBD method to multiple frames using the Richardson-Lucy algorithm under a maximum-likelihood (ML) framework. Similar ML approaches were

proposed by Sheppard *et al.* [32], Rajagopalan and Chaudhuri [29], Harikumar and Bresler [16] and others. The simulated annealing algorithm by McCallum [23] also estimates both the blur and image values simultaneously. This method tries to find a global minimum of a cost function by randomly perturbing blur and image estimates. Kundur and Hatzinakos [19] proposed an iterative approach based on recursive inverse filtering using nonnegativity and support constraints. They used conjugate gradient to minimize the associated cost function. In addition, several methods identify the blurring process by using special features such as edges and points, in the blurred image [28].

Other approaches have simplified the identification problem by parametrizing the point spread function (PSF). With some knowledge of the imaging system and environment, we can impose a blur degradation model with a few free parameters. Blur identification is then reduced to finding best estimates for these parameters. We generalize this approach to our blind superresolution problem. We enforce a parametric model upon the blurring process so that equation (1) becomes

$$\begin{aligned} \mathbf{f}_k &= DC(\sigma)E_k\mathbf{x} + \mathbf{n}_k, \quad 1 \le k \le p \qquad (6) \\ &= H_k(\sigma)\mathbf{x} + \mathbf{n}_k. \qquad (7) \end{aligned}$$

where the blurring operator $C$ is generated from a parameter set $\sigma$. The least squares solution of (7) is the minimizer to

$$\text{argmin}_{\mathbf{x}} \quad \sum_{k=1}^{p} \|\mathbf{f}_k - H_k(\sigma)\mathbf{x}\|_2^2 + \lambda \mathbf{x}^T Q \mathbf{x} \qquad (8)$$

where $\lambda$ controls the smoothness of the solution, and the stabilization (or regularization) matrix $Q$ is some symmetric positive definite matrix. For simplicity, we will consider $Q$ to be the identity matrix in this paper.[1] The minimizer to (8) can then

[1] We note that since any positive definite choice of $Q$ can be reduced to the standard case ($Q = I$) by redefining the variables, this assumption results in no loss of generality.

be expressed as follows:

$$\mathbf{x}(\sigma, \lambda) = (H(\sigma)^T H(\sigma) + \lambda I)^{-1} H(\sigma)^T \mathbf{f} \tag{9}$$

$$\mathbf{f} = \begin{bmatrix} \mathbf{f}_1 \\ \vdots \\ \mathbf{f}_p \end{bmatrix}, \quad H = \begin{bmatrix} H_1 \\ \vdots \\ H_p \end{bmatrix}. \tag{10}$$

Our approach to blind superresolution first estimates the unknown PSF parameter set $\sigma$ from raw data. Once an estimate of the PSF is available, the preconditioned conjugate gradient algorithm described in a previous paper by the authors [27, 26] is used to solve the non-blind superresolution problem. In the next section, we describe our parametrized blur estimation technique using generalized cross-validation.

## 5 Cross-Validation

Generalized cross-validation is a popular method for computing the regularization parameter [10]. In [27], we derived the formula for GCV for underdetermined linear systems. Not surprisingly, it has the same form as in the overdetermined case.

$$\lambda_{GCV} = \text{argmin}_\lambda \frac{\|(HH^T + \lambda I)^{-1}\mathbf{f}\|_2}{\text{tr}((HH^T + \lambda I)^{-1})}. \tag{11}$$

Reeves and Mersereau have used GCV for blur identification under an autoregressive moving average (ARMA) model [30]. In a recent study by Chardon, Vozel, and Chehdi [7], GCV has been shown to be an effective tool in parametric blur estimation. Motivated by these successes, we apply GCV to estimate both the PSF and regularization parameters for blind superresolution:

$$\{\sigma_{GCV}, \lambda_{GCV}\} = \text{argmin}_{\{\sigma,\lambda\}} \frac{\|(H(\sigma)H(\sigma)^T + \lambda I)^{-1}\mathbf{f}\|_2}{\text{tr}((H(\sigma)H(\sigma)^T + \lambda I)^{-1})}. \tag{12}$$

In [30], Reeves and Mersereau greatly simplified the minimization problem above by assuming the system matrix $H(\sigma)$ to be square and circulant, and hence, diagonalizable by the discrete Fourier transform. As a result, the numerator and denominator

of the objective function in (12) could be computed easily. However, this approach is not valid for blind superresolution because the system matrix will not typically be square or circulant. Without these assumptions, the numerator and denominator of (12) are prohibitively expensive to evaluate directly. In the later sections, we describe techniques to bound the objective function (12) efficiently and accurately.

Reeves and Mersereau simultaneously estimated the optimal blur and regularization parameters while keeping the image model parameters fixed. We found, however, that by setting the regularization parameter to some small number, the PSF parameters can be better estimated even in the presence of noise and a small number of frames. We then use the computed PSF to determine the appropriate regularization parameter based upon the data. Our intuition is that with under-regularization, the noise effect is exacerbated and moves the GCV criterion away from possible local minima. Furthermore, the estimated PSF is less biased away from the actual PSF even though the variance of the estimates can be larger. In what follows, we first use a small value of $\lambda = \lambda_0$, so that the estimated PSF parameters can be found by solving a one-dimensional nonlinear optimization problem

$$\sigma_{GCV} = \text{argmin}_\sigma \frac{\|(H(\sigma)H(\sigma)^T + \lambda_0 I)^{-1}\mathbf{f}\|_2}{\text{tr}((H(\sigma)H(\sigma)^T + \lambda_0 I)^{-1})}. \tag{13}$$

In the simplest case, the parameter set $\sigma$ consists of one parameter describing the smoothness of the blur, e.g. the standard deviation of a Gaussian PSF or the radius of a pillbox (out-of-focus) blur. Once a blur estimate $\hat{\sigma}$ is available, we compute the regularization parameter from

$$\hat{\lambda} = \text{argmin}_\lambda \frac{\|(H(\hat{\sigma})H(\hat{\sigma})^T + \lambda I)^{-1}\mathbf{b}\|_2}{\text{tr}((H(\hat{\sigma})H(\hat{\sigma})^T + \lambda I)^{-1})}. \tag{14}$$

## 6 Quadrature Rules with Lanczos Algorithm

For large systems, numerators and denominators in (13) and (14) are very expensive to evaluate directly. We first approximate the denominator using an unbiased trace estimator introduced by Hutchinson [17]: Let $U$ be a discrete random variable which takes the values -1 and +1 each with probability $\frac{1}{2}$, and let $\mathbf{u}$ be a vector whose entries are independent samples from $U$. Then the term $\mathbf{u}^T(HH^T + \lambda I)^{-1}\mathbf{u}$ is an unbiased estimator of $tr((HH^T + \lambda I)^{-1})$.

Now, in order to estimate both the numerators and denominators in (13) and (14), we need to estimate quadratic forms $\mathbf{v}^T f(M)\mathbf{v}$, where $M$ is some symmetric positive definite matrix and $f(\xi) = \xi^{-p},\ p = 1, 2$. There is extensive literature on the application of Gauss quadrature rules to bound bilinear forms; see papers by Golub and collaborators [3, 9, 11, 12, 14]. This paper applies these techniques to our blur identification problem.

Let the eigendecomposition of $M$ be given by $M = Q^T \Xi Q$, where $Q$ is an orthogonal matrix and $\Xi$ is a diagonal matrix of eigenvalues in increasing order. Then

$$\begin{aligned} \mathbf{v}^T f(M)\mathbf{v} &= \mathbf{v}^T Q^T f(\Xi) Q\mathbf{v} \\ &= \tilde{\mathbf{v}}^T f(\Xi)\tilde{\mathbf{v}} \\ &= \textstyle\sum_{i=1}^{n} f(\xi_i)\tilde{v}_i^2, \end{aligned} \tag{15}$$

where $\tilde{\mathbf{v}} = (\tilde{v}_i) \equiv Q\mathbf{v}$. Suppose that we have bounds on the spectrum of $M$, (e.g. by Gershgorin circle theorem), such that $a \leq \xi_1 \leq ... \leq \xi_n \leq b$. The last sum can be considered as a Riemann-Stieltjes integral with nondecreasing piecewise constant measure [14]

$$\sum_{i=1}^{n} f(\xi_i)\tilde{v}_i^2 = \int_a^b f(\xi) d\mu(\xi), \tag{16}$$

where the measure $\mu(\xi)$ is defined as

$$\mu(\xi) = \begin{cases} 0, & \text{if } \xi < \xi_1 \\ \sum_{j=1}^{i} \tilde{v}_j^2, & \text{if } \xi_i \leq \xi < \xi_{i+1} \\ \sum_{j=1}^{n} \tilde{v}_j^2, & \text{if } \xi_n \leq \xi. \end{cases} \tag{17}$$

The key to efficiently and accurately estimating the quadratic forms is that we can approximate the Riemann-Stieltjes integral (16) with Gauss-type quadrature rules.

The general form for quadrature rules is

$$\begin{aligned} I[f] &= \sum_{i=1}^{k} \omega_i f(\theta_i) + \sum_{j=1}^{l} \nu_j f(\tau_j), \\ \int_a^b f(\xi) d\mu(\xi) &= I[f] + R[f], \end{aligned} \tag{18}$$

where the weights $\omega_i$ and $\nu_j$ and the nodes $\theta_i$ are unknown, the nodes $\tau_j$ are predetermined, and $R[f]$ is the error term. The Gauss-type quadrature rules differ from one another by the number of prescribed nodes. If there are no prescribed nodes, then we obtain the standard Gauss quadrature:

$$I_G[f] = \sum_{i=1}^{k} \omega_i f(\theta_i). \tag{19}$$

If one node is prescribed, we get the Gauss-Radau quadrature rule; with two nodes prescribed, the Gauss-Lobatto rule:

$$I_R[f] = \sum_{i=1}^{k} \omega_i f(\theta_i) + \nu f(\tau), \tag{20}$$

$$I_L[f] = \sum_{i=1}^{k} \omega_i f(\theta_i) + \nu_1 f(\tau_1) + \nu_2 f(\tau_2). \tag{21}$$

The Gauss-Radau rule is often applied with either $\tau = a$ or $\tau = b$. Gauss-Lobatto has both endpoints prescribed, $\tau_1 = a, \tau_2 = b$. As will be described below, we can compute the unknown nodes $\theta_i$ and weights $\omega_i, \nu_i$ for these Gauss-type rules

from recurrence coefficients of sequences of orthogonal polynomials via the Lanczos bidiagonalization algorithm (cf. [11, 14]). The next subsection describes quadrature bounds on quadratic forms.

### 6.1 Quadrature Error and Bounds

The quadrature error $R[f]$ from (18) can be expressed as

$$R[f] = \frac{f^{(2k+l)}(\eta)}{(2k+l)!} \int_a^b \prod_{j=1}^{l} (\xi - \tau_j) \prod_{i=1}^{k} (\xi - \theta_i)^2 d\mu(\xi), \tag{22}$$

for some $\eta \in (a, b)$. We have the following bound for the Gauss quadrature rule [11].

**Theorem 1** *Suppose that* $f^{(2k)}(\xi) > 0,\ \forall k,\ \forall \xi,\ a < \xi < b,$ *then*

$$I_G[f] \leq \int_a^b f(\xi) d\mu(\xi). \tag{23}$$

**Proof.** For Gauss rule, there are no prescribed nodes, so

$$R_G[f] = \frac{f^{(2k)}(\eta)}{(2k)!} \int_a^b \prod_{i=1}^{k} (\xi - \theta_i)^2 d\mu(\xi). \tag{24}$$

Since $f^{(2k)}(\eta) > 0,\ \forall k,\ \forall \eta,\ a < \eta < b$, and $\int_a^b \prod_{i=1}^{k} (\xi - \theta_i)^2 d\mu(\xi) \geq 0$, then $R_G[f] \geq 0$. Therefore,

$$I_G[f] \leq \int_a^b f(\xi) d\mu(\xi). \tag{25}$$

Analogous theorems for Gauss-Radau and Gauss-Lobatto rules are also available:

**Theorem 2** *Suppose that* $f^{(2k+1)}(\xi) < 0,\ \forall k,\ \forall \xi,\ a < \xi < b,$ *then*

$$I_{Rb}[f] \leq \int_a^b f(\xi) d\mu(\xi) \leq I_{Ra}[f], \tag{26}$$

*where* $I_{Ra}$ $(I_{Rb})$ *corresponds to Gauss-Radau rule with prescribed node at* $\tau = a$ $(b)$.

**Proof.** Similar proof as in Theorem 1.

**Theorem 3** *Suppose that* $f^{(2k)}(\xi) > 0,\ \forall k,\ \forall \xi,\ a < \xi < b,$ *then*

$$\int_a^b f(\xi) d\mu(\xi) \leq I_L[f]. \tag{27}$$

*with prescribed nodes* $\tau_1 = a,\ \tau_2 = b.$

**Proof.** Similar proof as in Theorem 1.

Recall that the quadratic terms which we are interested in approximating have form $\mathbf{v}^T M^{-p} \mathbf{v}$, $p = 1, 2$. The function $f(\xi) = \xi^{-p}$ satisfies the hypotheses of Theorems 1,2, and 3, for $M$ positive definite ($a > 0$). Hence, we can apply the above theorems to bound $\int_a^b f(\xi) d\mu(\xi)$ with Gauss-type quadrature rules. Define

$$L[f] := \max(I_G[f], I_{Rb}[f]), \tag{28}$$

$$U[f] := \min(I_{Ra}[f], I_L[f]), \tag{29}$$

where $I_{Ra}$ ($I_{Rb}$) is the Gauss-Radau rule with the prescribed node at $\tau = a$ ($b$). We have the following bounds

$$L[f] \leq \int_a^b f(\xi) d\mu(\xi) \leq U[f]. \tag{30}$$

As the number of nodes $k$ increases the bounds $L[f]$ and $U[f]$ become tighter. To find quadrature bounds $L[f]$ and $U[f]$, we need the unknown weights and nodes $\omega_i$, $\nu_j$, and $\theta_i$. Appendix A describes how these quantities can be computed from sequences of orthogonal polynomials associated with the weight measure $d\mu(\xi)$. In practice, we use $U[f]$ as the approximate value for $\int_a^b f(\xi) d\mu(\xi) := \mathbf{v}^T f(M) \mathbf{v}$.

# 7 Experiments

We estimate blur and regularization parameters using the GCV criterion with quadrature rule bounds as described above. We use Matlab's CONSTR routine [18] to solve

the GCV minimization problem (13). For each function evaluation, we iterate with the Lanczos algorithm with the following stopping criterion:

$$\frac{U[f] - L[f]}{U[f]} \leq 0.01.$$

For our test image sequences, Lanczos usually terminates within 70 iterations, equivalent to 140 matrix-vector multiplies. The iteration count is quite low compared to the dimensions of the system matrix (usually in the tens of thousands).

Example I: In the first set of experiments, 16 LR frames are generated by blurring a $172 \times 172$ pixel HR image of the Stanford quad, with a $4 \times 4$ Gaussian blur and down-sampling by a factor of 4 in each dimension. We experiment with blurs of standard deviations 0.75, 1.0, and 2.0 and try to estimate these parameters assuming the support of the PSF to be known. We use the nonlinear PSF parameter estimation equation (13) with $\lambda_0 = 0.001$. In addition, we consider realizations of 60 dB, 30 dB SNR[2] and without additive Gaussian noise added to the LR frames. Tables 1 and 2 display the mean square error (MSE) of the PSF estimates under these conditions, with 10 randomly chosen LR frames and with all 16 frames, respectively. The formula for calculating percent MSE is [7]:

$$MSE(\hat{\sigma}) = 100 \frac{\|ch(\hat{\sigma}) - h(\sigma)\|_2^2}{\|h(\sigma)\|_2^2},$$

where $h(\hat{\sigma})$ is the PSF estimate of $h(\sigma)$, and the scaling factor

$$c = \frac{h(\hat{\sigma}) \cdot h(\sigma)}{\|h(\hat{\sigma})\|_2^2}.$$

These tables show excellent PSF reconstruction results for all cases. Next, in Figure 4, we compare quantitatively between blind superresolution reconstruction result against the original HR image. The actual blur standard deviation in this example is 0.75. We add white noise to the 10 LR frames to realize an SNR of 30 dB.

[2]Signal to noise ratio (SNR) is defined as $10 \log_{10} \frac{\sigma^2}{\sigma_{\mathbf{n}}^2}$, where $\sigma^2$, $\sigma_{\mathbf{n}}^2$ are variances of a clean frame and noise, respectively.

We compute first the PSF estimate from (13) using $\lambda_0 = 0.001$ as above. With this estimated PSF, we obtain a regularization parameter by solving (14). Figure 4 shows the result of blind superresolution using computed PSF and regularization parameters.

In the second set of experiments, we run tests for a pillbox (defocussed) blur. The parameter to be estimated is the radius of the blur. Our experiments test for out-of-focus blurs with radii 2 and 5 using the LR Stanford frames. We plot GCV values for radius taking values from 1 to 10 at 0.2 intervals in Figure 5. The plots show that GCV function achieves global minimum at the correct values in all cases.

Example II: The LR FLIR images in our final blind superresolution experiment are provided courtesy of Brian Yasuda and the FLIR research group in the Sensors Technology Branch, Wright Laboratory, WPAFB, OH. Each image is $64 \times 64$ pixels and a resolution enhancement factor of 5 is sought after. The objects in the scene are stationary, and 16 frames are acquired by controlled movements of a FLIR imager described in [15]. The frame to frame motions are accurately known for this sequence. We first estimate the blur variance assuming a Gaussian blur with support equaling the resolution enhancement factor, using (13) with $\lambda_0 = 0.001$. Similar to the process outline in example I, we compute the regularization parameter using the PSF estimate. Figure 6 shows a sample LR frame from the FLIR sequence and the resulting superresolved HR image using the computed variance and regularization parameter.

## 8 Conclusions

We propose a parametric blur and regularization estimation approach, based on the generalized cross-validation method, for blind restoration/superresolution. We solve a multivariate nonlinear minimization problem for these unknown parame-

ters. To efficiently and accurately estimate the numerator and denominator of the GCV objective function, we present Gauss-type quadrature techniques for bounding quadratic forms. Experimental results from a synthetic image sequence show that blur parameters are accurately approximated from Gaussian and pillbox blurs under various conditions. Image superresolution results using these computed values are visually pleasing as well. Successful experiments with a real FLIR image sequence illustrate that our techniques can be a foundation for data-driven efficient blind restoration/superresolution algorithms.

## References


[1] H. Andrews and B. Hunt. *Digital Image Restoration*. Prentice-Hall, Englewood Cliffs, NJ, 1977.

[2] G. Ayers and J. Dainty. Iterative blind deconvolution method and its applications. *Optics Letters*, 13(7), Jul. 1988.

[3] Z. Bai, M. Fahey, and G. Golub. Some large-scale matrix computation problems. *J. Comput. Appl. Math.*, 74:71–89, 1996.

[4] J. Biemond, R. Lagendijk, and R. Mersereau. Iterative methods for image deblurring. *Proc. IEEE*, 78(5):856–883, May 1990.

[5] D. Biggs and M. Andrews. Asymmetric iterative blind deconvolution of multiframe images. In F. Luk, editor, *SPIE Conference on Advanced Signal Processing Algorithms, Architectures, and Implementations VII*, volume SPIE Vol. 3461, San Diego, CA, Jul. 1998.

[6] L. Brown. A survey of image registration techniques. *ACM Computing Survey*, 24(4):325–376, Dec. 1992.

[7] S. Chardon, B. Vozel, and K. Chehdi. A comparative study between parametric blur estimation methods. In *ICASSP '99*, Phoenix, AZ, 1999.

[8] M. Elad. *Super-resolution Reconstruction of Images*. PhD thesis, The Technion - Israel Institute of Technology, Dec. 1996.

[9] G. Golub. Some modified matrix eigenvalue problem. *SIAM Review*, 15:318–334, 1973.

[10] G. Golub, M. Heath, and G. Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21:215–223, 1979.

[11] G. Golub and G. Meurant. Matrices, moments, and quadrature. In D. F. Griffiths and G. A. Watson, editors, *Numerical Analysis 1993*, pages 105–156. Longman, Essex, England, 1994.

[12] G. Golub and Z. Strakos. Estimates in quadratic formulas. *Numer. Algor.*, 8:241–268, 1994.

[13] G. Golub and U. von Matt. Generalized cross-validation for large-scale problems. *Journal of Computational and Graphical Statistics*, 6:1–34, 1997.

[14] G. Golub and J. Welsch. Calculation of Gauss quadrature rules. *Math. Comp.*, 23:221–230, 1969.

[15] R. Hardie, K. Barnard, and E. Armstrong. Joint MAP registration and high-resolution image estimation using a sequence of undersampled images. *IEEE Trans. on Image Processing*, 6(12), Dec. 1997.

[16] G. Harikumar and Y. Bresler. Perfect blind restoration of images blurred by multiple filters: Theory and efficient algorithms. *IEEE Trans. on Image Processing*, 8(2):202–219, Feb. 1999.

[17] M. Hutchinson. A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines. *Communications in Statistics, Simulation, and Computation*, 19:433–450, 1990.

[18] The Mathworks Inc. MATLAB, high-performance numeric computation and visualization software. Natick, MA, 1992.

[19] D. Kundur and D. Hatzinakos. Blind image deconvolution. *IEEE Signal Processing Magazine*, 13(3):43–64, May 1996.

[20] R. Lane and R. Bates. Automatic multidimensional deconvolution. *J. Opt. Soc. Am. A*, 4(1):180–188, Jan. 1987.

[21] W. Lukosz. Optical systems with resolving power exceeding the classical limit. *J. Opt. Soc. Am.*, 56(11):1463–1472, 1966.

[22] W. Lukosz. Optical systems with resolving power exceeding the classical limit II. *J. Opt. Soc. Am.*, 57(7):932–941, 1967.

[23] B. McCallum. Blind deconvolution by simulated annealing. *Optics Communications*, 75(2):101–105, Feb. 1983.

[24] N. Nguyen. *Numerical Techniques for Image Superresolution.* PhD thesis, Stanford University, expected Mar. 2000.

[25] N. Nguyen, P. Milanfar, and G. Golub. Blind superresolution with generalized cross-validation using Gauss-type quadrature rules. In *Proc. of the 33rd Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Oct. 1999.

[26] N. Nguyen, P. Milanfar, and G. Golub. A computationally efficient superresolution image reconstruction algorithm. Technical Report SCCM-99-04, Stanford University, Stanford, CA, 1999.

[27] N. Nguyen, P. Milanfar, and G. Golub. Preconditioners for regularized image superresolution. In *Proc. of ICASSP '99*, Phoenix, AZ, 1999.

[28] K. Nishi and S. Ando. Blind superresolving image recovery from blur-invariant edges. In *Proc. of ICASSP '94*, volume 5, pages V/85–88, 1994.

[29] A. Rajagopalan and S. Chaudhuri. A recursive algorithm for maximum likelihood-based identification of blur from multiple observations. *IEEE Trans. on Image Processing*, 7(7):1075–1079, Jul. 1998.

[30] S. Reeves and R. Mersereau. Blur identification by the method of generalized cross-validation. *IEEE Trans. on Image Processing*, 1(3):301–311, Jul. 1992.

[31] H. Shekarforoush and R. Chellappa. Data-driven multi-channel super-resolution with application to video sequences. *Journal of the Optical Society of America A*, 16(3):481–492, Mar. 1999.

[32] D. Sheppard, B. Hunt, and M. Marcellin. Iterative multiframe superresolution algorithms for atmospheric-turbulence-degraded imagery. *J. Opt. Soc. Am. A*, 15(4):978–992, Apr. 1998.

[33] G. Szegö. *Orthogonal Polynomials.* American Mathematical Society, 3rd edition, 1974.

[34] J. Woods and V. Ingle. Kalman filtering in two dimensions: Further results. *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-29:188–197, Apr. 1981.

# A Orthogonal Polynomials and the Lanczos Algorithm

Recall from Section 6 that we bound quadratic form $\mathbf{v}^T f(M)\mathbf{v}$, where $M$ is symmetric positive definite, with Gauss-type quadrature rules of the form

$$I[f] = \sum_{i=1}^{k} \omega_i f(\theta_i) + \sum_{j=1}^{l} \nu_j f(\tau_j).$$

As mentioned earlier, the unknown nodes $\theta_i$ and weights $\omega_i$, $\nu_i$ needed to calculate these quadrature values are related to the roots of orthogonal polynomials associated with the weight measure $d\mu(\xi)$ and the coefficients of their three-term recurrence. The following subsection describes orthogonal polynomials in more detail. Subsection A.2 outlines the Lanczos bidiagonlization used to efficiently compute the recurrence coeffcients and quadrature values.

## A.1 Orthogonal Polynomials

For a nondecreasing piecewise constant weight function $\mu(\xi)$, we can define a sequence of orthonormal polynomials $\{p_i\}_{i=0}^{n-1}$ such that

$$\int_a^b p_i(\xi)p_j(\xi)d\mu(\xi) = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases} \tag{31}$$

These polynomials satisfy a three-term recurrence relation (see [33]):

$$\xi \mathbf{p}_k(\xi) = T_k \mathbf{p}_k(\xi) + \beta_k p_k(\xi)\mathbf{e}_k, \quad k = 1, ..., n, \tag{32}$$

$$p_{-1}(\xi) = 0, \; p_0(\xi) = 1. \tag{33}$$

with

$$\mathbf{e}_k = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{p}_k = \begin{bmatrix} p_0(\xi) \\ \vdots \\ p_{k-1}(\xi) \end{bmatrix}, \quad T_k = \begin{bmatrix} \alpha_1 & \beta_1 & & \\ \beta_1 & \ddots & \ddots & \\ & \ddots & \alpha_{k-1} & \beta_{k-1} \\ & & \beta_{k-1} & \alpha_k \end{bmatrix}. \tag{34}$$

It can be shown that the nodes $\theta_i$ of Gauss quadrature rule are the eigenvalues of $T_k$, which are also the zeros of the polynomial $p_n$. The weights $\omega_i$ are the square of the first component of the normalized eigenvectors of $T_k$. The Gauss quadrature approximation is given by (see [13]):

$$I_G[f] = \|\mathbf{v}\|^2 \mathbf{u}_1^T f(T_k + \lambda I)\mathbf{u}_1, \tag{35}$$

where $\mathbf{u} = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T$ is a $k$-vector with one in the first entry and zeros elsewhere. For the Gauss-Radau rule, we need to adjust the last entry $\alpha_k$ of $T_k$ so that the adjusted tridiagonal matrix $\hat{T}_k$ has an eigenvalue at the prescribed node [9, 3]. For the Gauss-Lobatto rule, the last three nonzero entries $\beta_{k-1}, \alpha_k, \beta_{k-1}$ will be adjusted to prescribe eigenvalues at $a$ and $b$.

Following [3], for the Gauss-Radau rule, in order to prescribe a node at either $a$ or $b$, we replace the last entry $\alpha_k$ of $T_k$ by

$$\phi = \tau + s_{k-1}, \tag{36}$$

with $\tau = \{a, b\}$ and $s_{k-1}$ being the last component of the solution $\mathbf{s}$ to

$$(T_{k-1} - \tau I)\mathbf{s} = \beta_{k-1}^2 \mathbf{e}_{k-1}, \tag{37}$$

$$\beta_{k-1} = (\gamma_{k-1} + \gamma_k)\delta_{k-1}. \tag{38}$$

For Gauss-Lobatto rule, to prescribe nodes at $a$ and $b$, the last three nonzero entries $\beta_{k-1}, \alpha_k, \beta_{k-1}$ are replaced by $\psi, \phi, \psi$, where

$$\phi = \frac{s_{k-1}b - t_{k-1}a}{s_{k-1} - t_{k-1}}, \quad \psi^2 = \frac{b - a}{s_{k-1} - t_{k-1}}, \tag{39}$$

where $s_{k-1}$ and $t_{k-1}$ are the last components of the solutions $\mathbf{s}$ and $\mathbf{t}$ to

$$(T_{k-1} - aI)\mathbf{s} = \mathbf{e}_{k-1} \quad (T_{k-1} - bI)\mathbf{t} = \mathbf{e}_{k-1}. \tag{40}$$

For Gauss-Radau and Gauss-Lobatto rules, the quadrature approximations have the form

$$I[f] = \|\mathbf{v}\|^2 \mathbf{u}_1^T f(\tilde{T}_k + \lambda I)\mathbf{u}_1 \tag{41}$$

where $\tilde{T}_k$ is the tridiagonal matrix $T_k$ with the last few entries adjusted accordingly.

The entries of $T_k$ can be computed via the Lanczos bidiagonalization algorithm as follows.

## A.2 Lanczos Bidiagonalization

The matrix $T_k$ is the tridiagonal matrix resulting from $k$ iterations of Lanczos bidiagonalization algorithm for $M$ with $\frac{\mathbf{v}}{\|\mathbf{v}\|_2}$ as the starting vector. In Table 3, we include the Lanczos bidiagonalization factorization algorithm to efficiently compute the entries of $T_k$. After $k$ iterations of the Lanczos bidiagonalization algorithm, we obtain an orthogonal $m \times k$ matrix

$$Q_k = [\mathbf{q}_0 \cdots \mathbf{q}_{k-1}] \tag{42}$$

and an $n \times k$ orthogonal matrix

$$W_k = [\mathbf{w}_0 \cdots \mathbf{w}_{k-1}] \tag{43}$$

with the following relations

$$AW_k = Q_k B_k + \delta_k \mathbf{q}_k \mathbf{e}_k^T, \tag{44}$$

$$A^T Q_k = W_k B_k^T, \tag{45}$$

where $B_k$ is a $k \times k$ lower bidiagonal matrix

$$B_k = \begin{bmatrix} \gamma_1 & & & \\ \delta_1 & \ddots & & \\ & \ddots & \ddots & \\ & & \delta_{k-1} & \gamma_k \end{bmatrix} \tag{46}$$

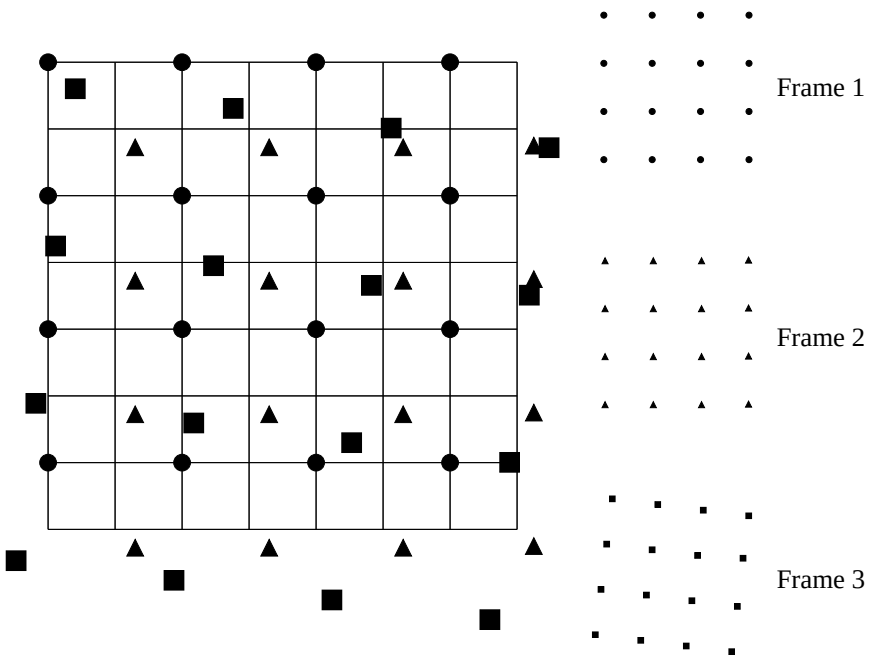


Figure 1: Low-resolution data on a high-resolution grid.

Combining (44) and (45) we get

$$AA^T Q_k = Q_k B_k B_k^T + \gamma_k \delta_k \mathbf{q}_k \mathbf{e}_k^T. \tag{47}$$

The tridiagonal $T_k$ in (32) is exactly $B_k B_k^T$.

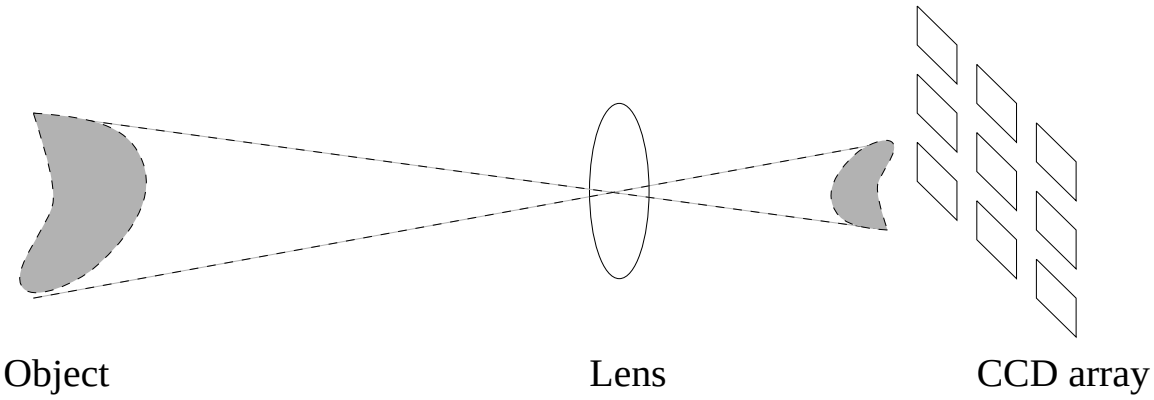


Figure 2: CCD camera model.

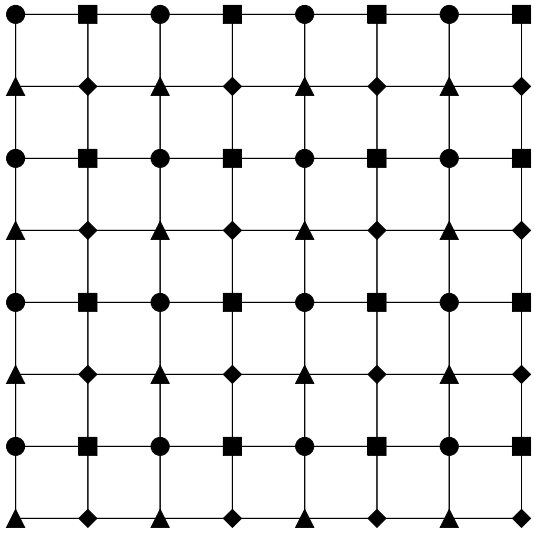

Figure 3: Image restoration as a special case of image superresolution.

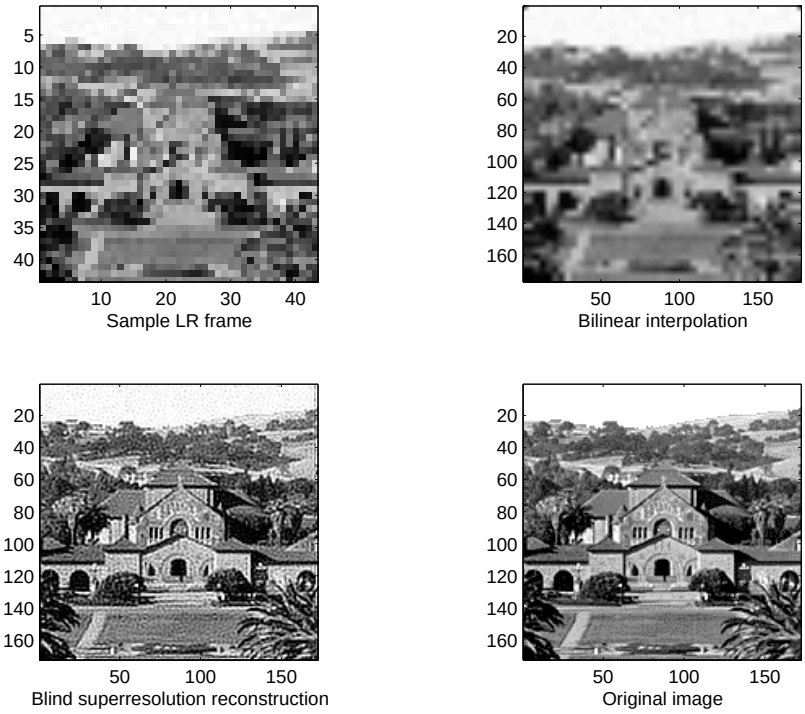


Figure 4: Blind superresolution for synthetic sequence - example I, 30dB.

| | SNR | | |
|---|---|---|---|
| $\sigma$ | $\infty$ dB | 60 dB | 30 dB |
| 0.75 | 3.43 | 3.27 | 4.79 |
| 1.0 | 0.02 | 0.02 | 0.53 |
| 2.0 | 0.01 | 0.01 | 0.11 |

Table 1: % MSE in PSF estimates for Gaussian blur with 10 randomly chosen frames - example I.

| | SNR | | |
|---|---|---|---|
| $\sigma$ | $\infty$ dB | 60 dB | 30 dB |
| 0.75 | 0.00 | 0.00 | 2.93 |
| 1.0 | 0.04 | 0.04 | 0.17 |
| 2.0 | 0.00 | 0.00 | 0.09 |

Table 2: % MSE in PSF estimates for Gaussian blur with all 16 frames available - example I

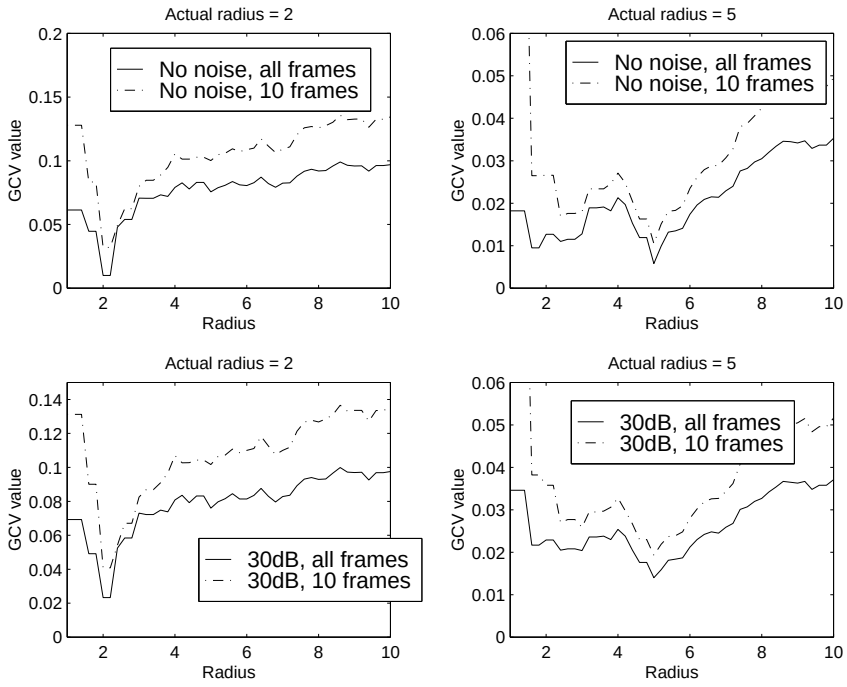


Figure 5: GCV plot for pillbox blur - example I.

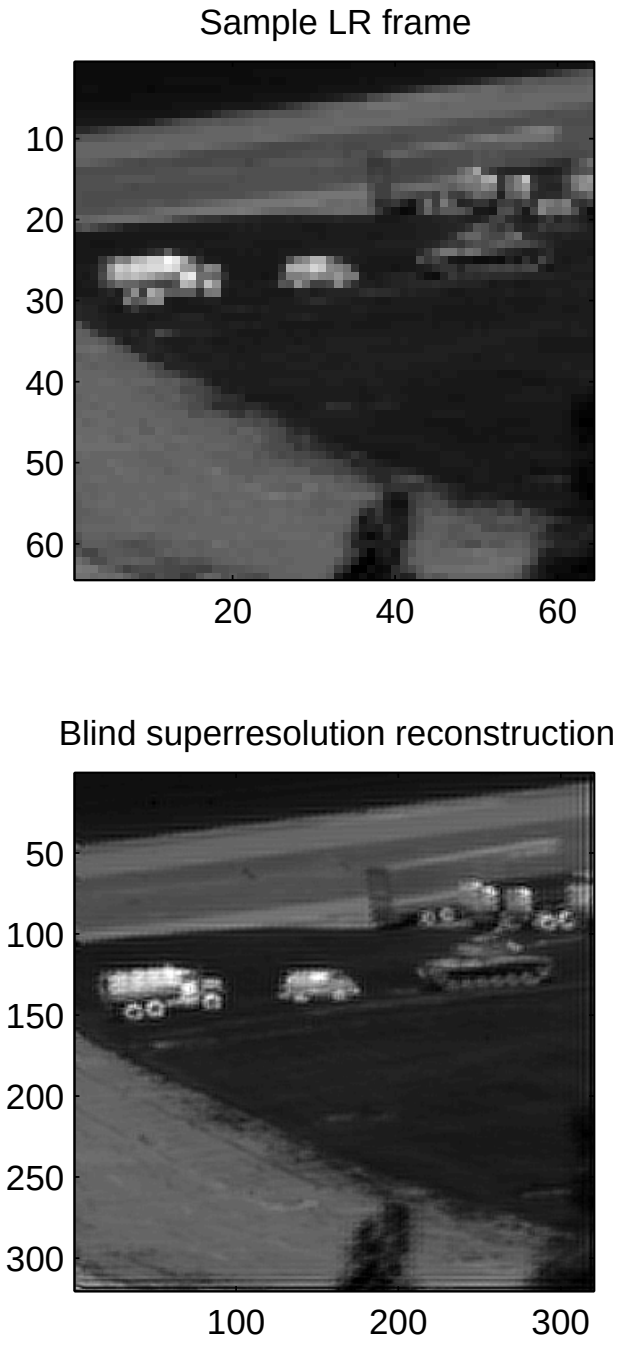


Figure 6: Blind superresolution for FLIR sequence - example II.

$\mathbf{q}_0 = \mathbf{b}/\|\mathbf{b}\|_2$

$\mathbf{s}_0 = A^T\mathbf{q}_0$

$\gamma_1 = \|\mathbf{s}_0\|_2$

$\mathbf{w}_0 = \mathbf{s}_0/\gamma_1$

**for** $k = 2 : n$ **do**

$\mathbf{r}_{k-1} = A\mathbf{w}_{k-2} - \gamma_{k-1}\mathbf{q}_{k-2}$

$\delta_{k-1} = \|\mathbf{r}_{k-1}\|_2$

$\mathbf{q}_{k-1} = \mathbf{r}_{k-1}/\delta_{k-1}$

$\mathbf{s}_{k-1} = A^T\mathbf{q}_{k-1} - \delta_{k-1}\mathbf{w}_{k-2}$

$\gamma_k = \|\mathbf{s}_{k-1}\|_2$

$\mathbf{w}_{k-1} = \mathbf{s}_{k-1}/\gamma_k$

**end**

Table 3: Lanczos bidiagonalization algorithm.