FOCAL PLANES

Focal planes are two-dimensional arrays of detectors employed for image formation. Charged coupled device (CCD) arrays are used in the visible spectrum, and infrared focal plane arrays (IRFPAs) are employed to sense thermal radiation. CCD arrays are mainly composed of silicon-based detectors and readout circuits. For IRFPAs, silicon is still the choice for readout circuits; however, it is not an effective material for detecting infrared radiation. Other material with narrow bandgap such as mercury cadmium telluride is used. The use of different materials for sensing, multiplexing, and readout in a hybrid scheme poses challenging connection problems. Moreover, the focal plane has to be cooled down to cryogenic temperatures. This cooling requirement highly increases the cost and complexity of IRFPAs. Currently, monolithic fabrication techniques that do not require sophisticated cooling mechanisms are being developed, with the potential of increasing the yield and lowering the cost of IRFPAs significantly. Rapid progress has been made in manufacturing uncooled IRFPAs operating at TV frame rates in recent years. At present, more demanding tasks such as missile seeking still rely on the hybrid technology (1–3).

The thermal radiation sensed by current infrared detectors lies in three spectral bands: long wavelength infrared (8 μm to 20 μm), medium wavelength infrared (3 μm to 5 μm), and short wavelength infrared (1 μm to 3 μm). A large fraction of the thermal radiation from objects in the ambient temperatures range is contained in the long wavelength infrared range and a small fraction is contained in the medium waveband. Small temperature differences in the ambient scene are effectively detected in the long wavelength band and, to a lesser extent, in the medium waveband. The peak emissions from artificial sources are mainly contained in the medium waveband, which makes it an excellent medium for the detection of hot bodies against a cooler background in military applications (4).

Focal planes are used in a variety of military, astronomical, medical, and industrial applications. Depending on the application, system parameters such as weight, size, sensitivity, resolution, power dissipation, and cost are determined. Present-day focal planes may consist of a few hundred to milliions of detectors. They can also be operated in different modes. In a *staring focal plane,* one detector is assigned to each pixel of the field of view. In the *scanning* mode, the focal plane is moved systematically over the field of view. Staring focal planes have the advantage of increased sensitivity, whereas a larger field of view is covered by a scanning focal plane. A step-staring sensor effectively brings the two operational modes together by staring at part of its total field of view for a time and then stepping to another part and staring again.

NOISE AND UNCERTAINTIES

Spatial response nonuniformity is an important problem with the use of IRFPAs. It arises because individual detectors on the focal plane exhibit different response characteristics from those of its neighboring elements. The response characteristics are described by parameters such as offset and gain of the detector element. The mean response of the detector to zero input flux is called the dark current offset. Generally, it is unique to each detector and varies with focal plane temperature, illumination history, and time elapsed from startup. Furthermore, detectors have reduced sensitivity at the top of their dynamic range, which is also to be compensated. Two or more auxiliary point sources may be used to calibrate the offset, gain, and higher order nonuniformities. This calibration procedure has to be repeated during operation in high-performance systems. Techniques that do not require auxiliary point sources based on neural networks are also being developed.

A second source of imperfection is the *electrical crosstalk* between detectors in close proximity. Crosstalk is measured by illuminating a detector element by a spot source and recording the response of the neighboring detectors. During the measurements care must be taken to localize the spot source at the exact center of the detector cell and to ascertain that the effective area of the spot source is smaller than the detector width. Otherwise, the optical crosstalk resulting from the energy induced on the adjacent detector cells may be mistaken as the result of electrical crosstalk.

Another source of noise encountered primarily in defense applications is impulsive noise due to gamma radiation. It manifests itself as noise samples with very large magnitudes that are independent in both space and time. Techniques suggested for treating bad detector elements are effective to mitigate impulsive noise as well. In particular, the samples affected by impulsive noise may be discarded, replaced by a user value, or interpolated using neighboring detector element values or time samples.

In a static scene, sampling the detector cells on the staring focal plane yields multiple observations of the same point in the field of view. These samples can be averaged to improve the signal-to-noise ratio when the noise can be modeled as additive Gaussian noise. For samples with impulsive noise components, robust estimation techniques such as the median and trimmed mean estimates should be used. For the scanning focal plane, the objects in the scene will move along the scan direction on the focal plane in a predictable manner. The same noise rejection techniques may be applied once the movement of the objects on the focal plane in the scan direction is accounted for. This approach is commonly known as *time delay integration* (TDI).

Focal Plane Signal Processing

At the signal-processing stage, one or more of the following tasks may be performed:

image enhancement,
noise rejection,
image restoration or reconstruction,
point source detection,
location, and
tracking.

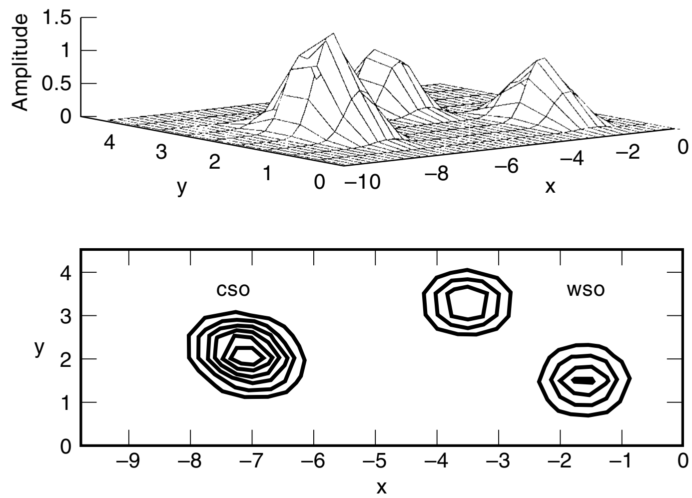

Figure 1. The response of the focal plane to two closely spaced objects and two widely separated objects.

The techniques that are used for image enhancement, reconstruction, and target tracking are covered elsewhere in this encyclopedia. We will mainly concentrate on point source processing techniques.

A *point source* within the context of infrared imaging is an object that is sufficiently far from the detector array so that it effectively acts as a point. Some examples of point sources may therefore be stars, missiles, and satellites. For a point source of a given intensity, the signal generated on a detector is determined by the *point spread function* of the optics, which is approximated well by the two-dimensional Gaussian function as well as by other design parameters such as detector shape, size, and the fill factor. We shall assume that the ideal detector cell response to a point source with unit amplitude is known by the user.

The conventional method for determining the location of a single point source in infrared imaging is *centroiding.* This method is based on finding the center of mass of a given frame and is applicable to cases where the point spread function is unknown. It is also useful in tracking problems where one is interested in tracking a group rather than individual point sources. Multiple widely separated sources can be located through the same means by partitioning the data into nonoverlapping segments. Matched filtering and track-before-detect methods provide useful means for separating widely spaced point sources. Alternative methods, however, are needed to resolve closely spaced sources, such as maximum-entropy Bayesian deblurring algorithms (e.g., see Refs. 20 and 21). Typical responses of the focal plane to two closely spaced objects (CSO) and widely separated objects (WSO) are depicted in Fig. 1.

MULTIPLE POINT SOURCE LOCATION PROBLEM

A set of p detector cells are located on the focal plane. Without loss of generality, the focal plane is taken to be the plane in three-dimensional space parallel to the xy axis and passing through the origin. The coordinates of points on this plane are therefore given by $(x, y, 0)$. In the problem under consideration, the radiation emitted by a number of point sources in three-dimensional space is intercepted by the detector cells.

A static point source with amplitude a_1 located at the z-axis with coordinates $(0, 0, z_1)$ produces a radiation density of $a_1\, s(x, y)$ at the output of the detector located at (x, y). In this representation $s(x, y)$ denotes the response of the detector cell to a unit amplitude point source. A static point source located at (x_1, y_1, z_1) produces the spatial shifted radiation density $a_1\, s(x - x_1, y - y_1)$ on the same detector.

In the simplest case, the radiation emitted by a single point source located at the fixed point (x_1, y_1, z_1) is measured by p detector cells on the focal plane. In accordance with the modeling assumptions made above, the individual detector cell responses are specified by the constant function of time

$$d_k(t) = a_1 s_k(x_1, y_1) \quad \text{for} \quad 1 \le k \le p \tag{1}$$

The dependency of this response term on the point source amplitude a_1 and location (x_1, y_1) has been made explicit, while its dependency on the position of the detector cell is implicitly recognized through the subscript k. In our modeling, each detector cell may have a distinctly different shape, although in many applications, the detector cells will have identical shapes. To employ concepts from contemporary array signal processing, the set of detector cell responses at time t shall be compactly represented by the $p \times 1$ detector cell response vector

$$\mathbf{d}(t) = [\, d_1(t)\, d_2(t)\, \ldots\, d_p(t)\,]^T \tag{2}$$

where $d_k(t)$ designates the response of the kth detector cell. We wish to use time-sampled values of the cell response vector (2) to estimate the locations of the point sources.

Upon substituting the cell response components of Eq. (1) into Eq. (2), an expression for the detector cell response vector is directly obtained. We shall express this detector cell response vector in the following form,

$$\begin{aligned} \mathbf{d}(t) &= a_1[\, s_1(x_1, y_1)\, s_2(x_1, y_1)\, \ldots\, s_p(x_1, y_1)\,]^T \\ &= a_1\, \mathbf{s}(x_1, y_1) \end{aligned} \tag{3}$$

The $p \times 1$ vector $\mathbf{s}(x_1, y_1)$ is referred to as the *steering vector* and it characterizes the manner in which the detector cells respond to a static point source located at (x_1, y_1, z_1). When there are multiple point sources irradiating the focal plane, the combined effect on the detector cell response is modeled as the sum of the responses generated by the individual point sources. Let there be m such point sources located at (x_l, y_l, z_l) for $1 \le l \le m$. The resultant detector cell response vector will be represented as a linear combination of steering vectors in the following compact notation:

$$\begin{aligned} \mathbf{d}(t) &= [\ \mathbf{s}(x_1, y_1) \vdots\ \mathbf{s}(x_2, y_2) \vdots\ \cdots\cdots\ \vdots \mathbf{s}(x_m, y_m)\ ] \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} \\ &= S(\theta)\mathbf{a} \end{aligned} \tag{4}$$

The $p \times m$ steering matrix $S(\boldsymbol{\theta})$ has the m steering vectors associated with the individual steering vectors as its columns and the $m \times 1$ multiple point source amplitude vector $\mathbf{a}$ has the individual point source amplitudes as its components. For notational brevity, the steering matrix $S(\boldsymbol{\theta})$ has been expressed as an explicit function of the $2m \times 1$ multiple point source location vector as specified by

$$\boldsymbol{\theta} = \begin{bmatrix} x_1 & y_1 & x_2 & y_2 & \cdots & x_m & y_m \end{bmatrix}^T \tag{5}$$

Equation (4) provides the idealized expression for the detector cell responses in the noise free case where m spatially stationary point sources radiate the infrared (IR) focal plane. In the more realistic case, the cell responses are corrupted by sensor noise and other extraneous influences. We shall quantify these extraneous factors by an additive $p \times 1$ "noise" vector $\mathbf{w}(t)$. To estimate the multiple point source location vector $\boldsymbol{\theta}$ and multiple point source amplitude vector $\mathbf{a}$, we shall use the following set of time samples of the noise-corrupted detector cell response vector

$$\mathbf{d}(t_n) = S(\boldsymbol{\theta})\mathbf{a} + \mathbf{w}(t_n) \quad \text{for} \quad 1 \le n \le N \tag{6}$$

where the time sampling scheme $\{t_n\}$ need not be uniform. Examination of these time samples reveals that they are each additive noise-corrupted measurements of the constant signal vector $S(\boldsymbol{\theta})\mathbf{a}$.

Least Squared Error Modeling

The task of multiple point source detection and location is basically that of using the N sampled values of the detector cell response vector [Eq. (6)] to estimate the $\mathbf{a}$ amplitude vector and the $\boldsymbol{\theta}$ multiple point source location vector. In this section, these parameters are selected so as to minimize the squared error criterion:

$$c(\mathbf{a}, \boldsymbol{\theta}) = \sum_{n=1}^{N} [\mathbf{d}(t_n) - S(\boldsymbol{\theta})\mathbf{a}]^T [\mathbf{d}(t_n) - S(\boldsymbol{\theta})\mathbf{a}] \tag{7}$$

The minimization of this squared error criterion is equivalent to the maximization of the likelihood function when the noise samples are temporally and spatially independent and identically distributed samples from a Gaussian distribution. When the samples from different detectors are either dependent or not identically distributed, a weighted squared error criterion can be used. Upon examination of the squared error criterion [Eq. (7)], the multiple point source amplitude vector $\mathbf{a}$ is seen to enter in a quadratic manner while the multiple point source location vector $\boldsymbol{\theta}$ appears in a highly nonlinear fashion. This being the case, a closed form expression for an optimum selection of $(\mathbf{a}, \boldsymbol{\theta})$ that minimizes this criterion does not exist. One must therefore appeal to nonlinear optimization methods. The computational complexity of these methods is a function of the number of variables in the minimization problem. Fortunately, as will be shown shortly, it is possible to separate the selections of the multiple point source amplitude vector $\mathbf{a}$ from the multiple source location vector $\boldsymbol{\theta}$ by exploiting the quadratic manner in which the former enters the criterion (7). This separation significantly decreases the computational complexity.

The convergence rate of the nonlinear programming algorithm is affected by the structure of the steering matrix $S(\boldsymbol{\theta})$.

In practice, faster convergence rates are achieved when $S(\boldsymbol{\theta})$ is decomposed as the product of a $p \times m$ matrix $Q(\boldsymbol{\theta})$, whose orthonormal column vectors span the range space of $S(\boldsymbol{\theta})$, and a $m \times m$ nonsingular upper triangular matrix $R(\boldsymbol{\theta})$, that is

$$S(\boldsymbol{\theta}) = Q(\boldsymbol{\theta})R(\boldsymbol{\theta}) \tag{8}$$

where

$$Q(\boldsymbol{\theta})^T Q(\boldsymbol{\theta}) = I_m \tag{9}$$

in which I_m is the $m \times m$ identity matrix. This QR decomposition can be achieved by applying the Gram–Schmidt orthogonalization procedure to the full rank matrix $S(\boldsymbol{\theta})$. Substituting Equations (8) and (9) into the squared error criterion of Eq. (7) and its minimization with respect to the multiple point source amplitude vector $\mathbf{a}$ yields the optimum multiple point source amplitude vector $\mathbf{a}^o$ as

$$\mathbf{a}^o = R(\boldsymbol{\theta})^{-1} Q(\boldsymbol{\theta})^T \hat{\mathbf{d}} \tag{10}$$

where

$$\hat{\mathbf{d}} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{d}(t_n) \tag{11}$$

Moreover, the value of the squared error criterion of Eq. (7) for this optimum choice is given by

$$c(\mathbf{a}^o, \boldsymbol{\theta}) = \sum_{n=1}^{N} \mathbf{d}(t_n)^T \mathbf{d}(t_n) - N\hat{\mathbf{d}}^T Q(\boldsymbol{\theta})Q(\boldsymbol{\theta})^T \hat{\mathbf{d}} \tag{12}$$

An examination of Eqs. (10) and (12) reveals that the optimum selections of the multiple point source location and amplitude vectors have been decoupled. In principle, one first finds the multiple point source location vector $\boldsymbol{\theta}$ that minimizes criterion [Eq. (12)]. Once this vector has been obtained, it is substituted into Eq. (10) to obtain the corresponding optimum multiple source amplitude vector. The minimization problem of Eq. (12) is highly nonlinear in the unknown multiple point source location vector $\boldsymbol{\theta}$. Nonlinear programming techniques must therefore be used to achieve this minimization.

Nonlinear Programming Solution

Many nonlinear programming techniques are based on the principal of incrementally perturbing the parameters to be optimized so that the functional being minimized takes on monotonically decreasing values. Various nonlinear programming algorithms are distinguished by the manner in which the perturbation vector and step size scalar are chosen (6–7). We shall employ the Gauss-Newton method, whose perturbation vector is specified by

$$\boldsymbol{\delta}_k^{(GN)} = -[J(\boldsymbol{\theta}_k)^T J(\boldsymbol{\theta}_k)]^{-1} J(\boldsymbol{\theta}_k)^T e(\mathbf{a}^o, \boldsymbol{\theta}_k) \tag{13}$$

in which the residual error vector $\mathbf{e}(\mathbf{a}^o, \boldsymbol{\theta}_k)$ is given by

$$\mathbf{e}(\mathbf{a}^o, \boldsymbol{\theta}) = (I - Q(\boldsymbol{\theta})Q(\boldsymbol{\theta})^T)\,\hat{\mathbf{d}} \tag{14}$$

and the $J(\boldsymbol{\theta}_k)$ is the Jacobian matrix. Closed form expressions for the elements of the Jacobian matrix are given in Ref. 8.

The effectiveness of descent algorithms such as the Gauss-Newton method largely depends on the initial choice of the composite location parameter vector. If a poor initial point is chosen, any descent algorithm may generally converge to a poor relative minimum. The sequential orthogonal projection algorithm is also called a coordinate descent algorithm (7) and has proven to be a useful initial point selection procedure in various applications (9–13). This procedure is based on sequentially increasing the number of point sources in the model, and with each new source added, an estimate of the location for that source is made using a direct search method.

OUTLIER DETECTION

The widespread use of the least squared error criterion is justified by its equivalence to the maximum likelihood criterion for independent identically distributed Gaussian noise. It furthermore provides mathematical tractability. Unfortunately, estimates obtained through LSE criterion are not asymptotically efficient when the noise is non-Gaussian. Symmetric non-Gaussian noise typically gives rise to estimates with high variance, whereas the estimates obtained in the presence of asymmetric non-Gaussian noise are biased as well.

A widely accepted approach to cleanse the input data from such outlying samples is to detect and remove aberrant data points before carrying out the location estimation phase. This can be achieved by using robust time delay integration techniques to a limited extent when the impulsive behavior is not severe. However, this approach may not be effective in the vicinity of a point source, especially if the amplitude of the point source is unknown. This being the case, we describe an outlier detection scheme that can be employed after the location and amplitude estimation step. Once the outliers are detected, they will be discarded and the estimation procedure repeated with the remaining samples. The iterations will be continued until no outlying samples are detected. Different outlier detection techniques are described in Ref. 14.

As the outlier samples are discarded, the sizes of the detector cell response vectors $\mathbf{d}(t_n)$ may not be identical for all t_n. Consequently, Eq. (10) cannot be used to estimate the amplitudes. To accommodate missing data, the model in Eq. (6) has to be modified. The data are represented by a composite detector cell response vector, which is obtained by concatenating the available detector cell response vectors. This gives an equivalent representation covering the cases of missing data as well as the case in which some of the point sources are moving. When the composite detector cell response and noise vector $\mathbf{d}$ and $\mathbf{w}$ are constructed as

$$\mathbf{d} = \begin{bmatrix} \mathbf{d}(t_1) \\ \mathbf{d}(t_2) \\ \vdots \\ \mathbf{d}(t_N) \end{bmatrix} \qquad \mathbf{w} = \begin{bmatrix} \mathbf{w}(t_1) \\ \mathbf{w}(t_2) \\ \vdots \\ \mathbf{w}(t_N) \end{bmatrix} \tag{15}$$

the data may be compactly represented as

$$\mathbf{d} = S(\boldsymbol{\theta})\mathbf{a} + \mathbf{w} \tag{16}$$

where $S(\boldsymbol{\theta})$ denotes composite steering matrix

$$S(\boldsymbol{\theta}) = \begin{bmatrix} S(\boldsymbol{\theta}; t_1) \\ S(\boldsymbol{\theta}; t_2) \\ \vdots \\ S(\boldsymbol{\theta}; t_N) \end{bmatrix} \tag{17}$$

where the dependence of the steering matrices on the time samples indicates their possible dependence on time when the sources are moving.

After the location estimation procedure is completed, the estimated parameters $\mathbf{a}^o$ and θ^o may be employed to obtain the "residual error vector" $\mathbf{e}$ as specified by

$$\mathbf{e} = \mathbf{d} - S(\boldsymbol{\theta}^o)\mathbf{a}^o \tag{18}$$

The elements of the residual error vector therefore indicate the mismatch between the actual and the estimated responses of the detector cells. Substituting the expression for the detector cell response vector of Eq. (16), the representation of the residual error vector is given by

$$\mathbf{e} = [S(\boldsymbol{\theta})\mathbf{a} - S(\boldsymbol{\theta}^o)\mathbf{a}^o] + \mathbf{w} \tag{19}$$

The first term in brackets in Eq. (19) vanishes if the estimates of the amplitude and location parameter vectors are identical to their actual values, that is, $\mathbf{a}^o = \mathbf{a}$ and $\boldsymbol{\theta}^o = \boldsymbol{\theta}$. This ideal situation is rarely achieved with a finite sample size, but may be approximated for symmetric distributions if the sample size is sufficiently large. For the mathematical analysis, we will assume that the actual and estimated values of amplitude and location parameter vectors are close enough so that the residual error vector is dominated by its input noise vector term $\mathbf{w}$.

In our modeling, the elements of the $L \times 1$ input noise vector $\mathbf{w}$ are realizations of the random variable w whose probability density function (pdf) is designated by p_w. This random variable is in turn generated as the sum of two independent random variables w_g and w_γ so that

$$w = w_g + w_\gamma \tag{20}$$

In many applications, w_g is modeled as a zero-mean Gaussian random variable with variance σ_g^2. The second random variable w_γ is usually non-Gaussian with nonzero mean and/or has a variance higher than σ_g^2. The scenario may be further complicated if some samples of w have no contribution from w_γ. This will be the case if samples of w_γ are generated from

$$w_\gamma = \begin{cases} \gamma & \text{with probability } \epsilon \\ 0 & \text{with probability } (1-\epsilon) \end{cases} \tag{21}$$

where γ is a sample from the so-called "contaminating" distribution with pdf p_γ and ϵ is from the closed interval [0, 1]. Such a random variable γ could represent impulsive noise encountered in infrared estimation problems. The relationship between the probability density functions of w_γ and γ is given by

$$p_{w_\gamma}(w_\gamma) = (1-\epsilon)\,\delta(w_\gamma) + \epsilon\, p_\gamma(w_\gamma) \tag{22}$$

where p_γ designates the pdf of γ. The samples of w_γ which take on the value zero are generated by the $(1 - \epsilon)\,\delta(w_\gamma)$ term where $\delta(.)$ is the Dirac delta function. Since w is defined in Eq. (20) as the sum of two independent random variables w_g and w_γ, its probability density function is specified by the convolution integral

$$p_w(w) = \int_{-\infty}^{\infty} p_g(\eta)\, p_{w_\gamma}(w-\eta)\, d\eta \tag{23}$$

where p_g designates the probability density function of w_g. Then the probability density function of w is given by

$$p_w = (1-\epsilon)\, p_g + \epsilon \int_{-\infty}^{\infty} p_g(\eta)\, p_\gamma(w-\eta)\, d\eta \tag{24}$$

Moreover, the closed form expressions for the mean μ_w and variance σ_w^2 of the random variable w are given by

$$\mu_w = \mu_g + \epsilon\, \mu_\gamma \tag{25}$$

$$\sigma_w^2 = \sigma_g^2 + \epsilon\, \sigma_\gamma^2 + \epsilon\,(1-\epsilon)\, \mu_\gamma^2 \tag{26}$$

where the means and variances of w_g and γ are denoted by μ_g, μ_γ and σ_g^2, σ_γ^2, respectively.

Equations (25) and (26) indicate that when the contaminating distribution γ has zero mean, only the variance of w increases. When γ has a nonzero mean, a further increase of the variance is accompanied by a shift in the mean as well. Examination of p_w in Eq. (24) reveals that the samples that are significantly affected by γ are located at the tails of p_w. A general approach to detect such samples is to determine the samples of w whose magnitude is larger than a threshold. The test statistic for the ith sample of w, $\tau(w_i)$, is given by

$$\tau(w_i) = \frac{w_i - \mu_g}{\sigma_g} \tag{27}$$

The outlier detection scheme described in this section may be integrated into a point source location and amplitude estimation method so that each iteration consists of an amplitude and location estimation followed by a step of detection of the outlying points in the residual error vector $\mathbf{e}$ [see Eq. (18)]. After each outlier detection step, all points that are declared as outliers are removed, and a new iteration of amplitude and location estimation is initiated. This iterative procedure is repeated until no samples of the residual error vector contains an outlier for the estimated amplitudes and locations.

For the initial iterations, the estimates of $\mathbf{a}$ and $\boldsymbol{\theta}$ may not be sufficiently close to their actual values. In that case, the residuals are dominated by the errors in the estimates of $\mathbf{a}$ and $\boldsymbol{\theta}$ rather than the additive noise. Then even the residuals corresponding to the samples without impulsive noise may be larger than the given threshold. To safeguard against excessive rejection, we use a modified rejection rule

$$\text{reject } w_i \quad \text{if } w_i > 3 \bullet \max(\hat{\sigma}_w, \sigma_g) \tag{28}$$

where $\hat{\sigma}_w$ is the sample standard deviation.

For distributions with nonzero mean, the estimates of the amplitudes obtained through Eq. (10) are typically biased. In fact, if the sample mean of w converges to $\boldsymbol{\mu}_w$ in probability, then the estimated amplitude vector is given by

$$\mathbf{a}^o = \mathbf{a} + R(\boldsymbol{\theta})^{-1}Q(\boldsymbol{\theta})^T \boldsymbol{\mu}_{\boldsymbol{w}} \quad (29)$$

where $\boldsymbol{\mu}_w$ is a $L \times 1$ vector with elements $\boldsymbol{\mu}_w$. Since the expression for $\boldsymbol{\mu}_w$ is given in Eq. (20) as the sum of μ_g and $\epsilon\,\mu_\gamma$, the $L \times 1$ vector $\boldsymbol{\mu}_w$ is also specified by

$$\boldsymbol{\mu}_{\boldsymbol{w}} = \mu_g + \epsilon\,\mu_\gamma \quad (30)$$

Hence, compensation of the nonzero mean may be achieved by modifying Eq. (10) as

$$\mathbf{a}^o = R(\theta)^{-1}Q(\theta)^T\,(\mathbf{d} - \boldsymbol{\mu}_{\boldsymbol{w}}) \quad (31)$$

The mean of $\boldsymbol{w}$ has to be either known or estimated beforehand to achieve an unbiased estimate of $\mathbf{a}$. Fortunately, it can be generally estimated from the part of the data that does not include a point source.

SUBSPACE METHODS

In the following sections we will describe two other location-finding methods based on an eigendecomposition of the data. These methods are called *subspace methods* since they involve decomposing the data into their components in two subspaces. Subspace methods have received considerable attention in the 1980s, inspired by the work of Pisarenko and Schmidt (15–16). Unfortunately, this very powerful class of high-resolution algorithms are not directly applicable to infrared point source location problems. Since the amplitudes of point sources are constants in time, eigenanalysis of the correlation matrix of the detector cell vector typically does not reveal the number and locations of the point sources. Nevertheless, subspace methods can still be applied to point source location estimation problems if the response of the detector cells are separable in their location parameters, and they are expressed by a data matrix (frame) rather than a detector cell response vector. Then it is possible to express the response due to a point source as the outer product of two vectors where the first vector depends on the x, and the second vector depends on the y location of the point source only. In the presence of multiple point sources, the outer products corresponding to each source are superimposed.

The location estimation procedures exploit the fact that the principal singular vectors of the data matrix span the same space as the basis vectors forming the outer products, and the other singular vectors are orthogonal to these basis vectors. A procedure that is predicated on the first property is called a *signal subspace method* while a procedure based on the second property is called a *noise subspace method.* In the absence of noise, the number of point sources may be determined as the number of nonzero singular values of the data matrix. In the presence of noise, the number of singular values that are significantly larger than others may be chosen to model the data. As opposed to the formulation in the previous sections, where the detector output was expressed by vectors, a matrix formulation shall be adopted here. This formulation is not only more convenient, but is necessary to apply subspace concepts.

Problem Formulation for Subspace Methods

Let there be m point sources radiating the focal plane. The radiation density induced at the observation point $(x, y, 0)$ on the focal plane by the ith point source is given by $a_i\, s(x - x_i, y - y_i)$. In this expression, a_i, x_i and y_i are, respectively, the amplitude and the x and y coordinates of the ith point source, and $s(x, y)$ is the response induced by a unit amplitude point source located at $(0, 0, z_i)$.

In the LSE method, the locations of the elements of the detector array were arbitrary. In this section, however, it will be assumed that the detectors are placed in the focal plane on a rectangular grid. Moreover, the detector array is uniformly sampled along the x and y directions with sampling rates T_x and T_y such that N_x and N_y samples are obtained in each direction. For a unit amplitude point source located at (x_i, y_i, z_i), this set of data can be expressed in a $N_x \times N_y$ data matrix $S(x_i, y_i)$ whose (k, l) component is specified by

$$S(x_i, y_i)_{k,l} = s([k-1]T_x - x_i, [l-1]T_y - y_i) \quad (32)$$

The data matrix $S(\boldsymbol{\theta})$, defining the combined effect of the m sources, will be the weighted sum of the individual data matrices, that is,

$$S(\boldsymbol{\theta}) = \sum_{i=1}^{m} a_i\, S(x_i, y_i) \quad (33)$$

where $\boldsymbol{\theta}$ is the $2m \times 1$ unknown location parameter vector

$$\boldsymbol{\theta} = [x_1\, y_1\, x_2\, y_2\, \ldots\, x_m\, y_m]^T \quad (34)$$

in accordance with Eq. (5). For the purposes of this study, we will restrict ourselves to signals that are separable in their parameters. If the signal generated by a point source, $s(x, y)$, is separable in the variables x and y, then $s(x, y)$, can be expressed as the product of functions of x and y, that is,

$$s(x, y) = s_x(x)s_y(y) \text{ for all } x, y \in \Re \quad (35)$$

Since the sampling scheme is uniform on the focal plane, the data matrix generated by the ith source is readily shown to be the scalar multiple of the outer product of two steering vectors $\mathbf{s}_x(x_i)$ and $\mathbf{s}_y(y_i)$, that is,

$$S(x_i, y_i) = a_i\, \mathbf{s}_x(x_i)\, \mathbf{s}_y^T(y_i) \quad (36)$$

where

$$\mathbf{s}_x(x_i) = [s_x(-x_i)\; s_x(T_x - x_i)\; \ldots\; s_x((N_x - 1)T_x - x_i)]^T \quad (37)$$

and

$$\mathbf{s}_y(y_i) = [s_y(-y_i)\; s_y(T_y - y_i)\; \ldots\; s_y((N_y - 1)T_y - y_i)]^T \quad (38)$$

The total signal generated by m point sources is given by the data matrix $S(\boldsymbol{\theta})$ defined in Eq. (33). For separable signals, this expression may be conveniently expressed as

$$S(\boldsymbol{\theta}) = S_x(\mathbf{x})\;A\;S_y^T(\mathbf{y}) \quad (39)$$

where the $N_x \times m$ and $N_y \times m$ matrices $S_x(\mathbf{x})$ and $S_y(\mathbf{y})$ have the steering vectors $\mathbf{s}_x(x_i)$ and $\mathbf{s}_y^T(y_i)$ as their columns, that is,

$$S_x(\mathbf{x}) = [\ \mathbf{s}_x(x_1) \vdots\ \mathbf{s}_x(x_2) \vdots\ \cdots\cdots \vdots\ \mathbf{s}_x(x_m)\ ] \tag{40}$$

$$S_y(\mathbf{y}) = [\ \mathbf{s}_y(y_1) \vdots\ \mathbf{s}_y(y_2) \vdots\ \cdots\cdots \vdots\ \mathbf{s}_y(y_m)\ ] \tag{41}$$

The unknown amplitudes constitute the diagonal elements of the diagonal matrix A so that

$$A = \begin{bmatrix} a_1 & & & 0 \\ & a_2 & & \\ & & \ddots & \\ 0 & & & a_m \end{bmatrix} \tag{42}$$

and the vectors $\mathbf{x}$ and $\mathbf{y}$, appearing in Eqs. (39), (40), and (41), are the $m \times 1$ location parameter vectors

$$\mathbf{x} = [x_1\, x_2\, \ldots\, x_m]^T \quad \text{and} \quad \mathbf{y} = [y_1\, y_2\, \ldots\, y_m]^T \tag{43}$$

Observations.

The problem at hand is to estimate the m amplitudes a_i and the $2m \times 1$ parameter vector $\boldsymbol{\theta}$. This is equivalent to estimating the $m \times 1$ parameter vectors $\mathbf{x}$, $\mathbf{y}$ and the diagonal elements of the matrix A.

As Eq. (39) suggests, the estimation procedures for $\mathbf{x}$ and $\mathbf{y}$ would be identical, except that the estimation procedure for $\mathbf{x}$ would involve $S_x(\mathbf{x})$ and the estimation procedure for $\mathbf{y}$ would involve $S_y(\mathbf{y})$.

Since the parameters of a given point source are defined by a unique set of amplitude and x and y coordinates, once all the amplitudes a_i and location vectors $\mathbf{x}$ and $\mathbf{y}$ are obtained, they have to be paired so that the data are best described by the parameter set. For point sources with equal amplitudes, the number of possible parameter sets is $m!$ and the number grows to $(m!)^2$ for point source with unequal amplitudes. Evidently, the amount of computation required may be unacceptable for large values of m.

Eigenanalysis for Separable Frames

The estimation procedure will be complicated by additive noise. In particular, the observed data matrix D will be given by

$$D = S(\boldsymbol{\theta}) + W \tag{44}$$

where W is taken to be a $N_x \times N_y$ matrix with elements from a wide sense stationary random process. It is also assumed that the elements of W are zero mean and uncorrelated, that is, for any two elements w_{ij} and w_{kl} of W

$$E\{w_{ij}w_{kl}\} = \sigma^2\ \delta(i-k, j-l)$$

where 'E' denotes the expected value operator and δ is the Kronecker delta function.

We will call the $E\{DD^T\}$, the correlation matrix of $\mathbf{x}$ and will denote it by $R(\mathbf{x})$. Similarly, the $E\{D^TD\}$ will be called correlation matrix of $\mathbf{y}$ and will be denoted by $R(\mathbf{y})$. The reason for the appearance of $\mathbf{x}$ and $\mathbf{y}$ will be clear shortly. The closed form expression for $R(\mathbf{x})$ is given by

$$\begin{aligned} R(\mathbf{x}) &= E\{DD^T\} \\ &= E\{(S(\boldsymbol{\theta})+W)(S(\boldsymbol{\theta})+W)^T\} \\ &= E\{S(\theta)S(\boldsymbol{\theta})^T\} \\ &\quad + E\{S(\boldsymbol{\theta})W^T\} + E\{WS(\boldsymbol{\theta})^T\} + E\{WW^T\} \\ &= S(\boldsymbol{\theta})S(\boldsymbol{\theta})^T + N_y\sigma^2 I_{N_x} \end{aligned} \tag{45}$$

where I_{N_x} is the $N_x \times N_x$ identity matrix. Similarly,

$$\begin{aligned} R(\mathbf{y}) &= E\{D^TD\} \\ &= S(\theta)^T S(\theta) + N_x\sigma^2 I_{N_y} \end{aligned} \tag{46}$$

Our objective is to develop techniques for determining the number and locations of the point sources based on the eigendecomposition of the data matrix. The following theorem provides a means for achieving this objective.

Theorem 1. Let the rank of $S_x(\mathbf{x})$ and $S_y(\mathbf{y})$ each be m. Furthermore, let $\lambda_1, \lambda_2, \ldots, \lambda_{N_x}$ be the eigenvalues of $R(\mathbf{x})$ and $\nu_1, \nu_2, \ldots, \nu_{N_x}$ be the eigenvalues of $S(\boldsymbol{\theta})S(\boldsymbol{\theta})^T$. Then

$$\lambda_i = \begin{cases} \nu_i + N_y\sigma^2 & \mathit{for}\ i = 1, 2, \ldots, m \\ N_y\sigma^2 & \mathit{for}\ i = m+1, m+2, \ldots, N_x \end{cases} \tag{47}$$

Furthermore, if $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{N_x}$ are the corresponding eigenvectors of $R(\mathbf{x})$,

$$\mathbf{u_i} \in \mathit{Range}\,\{S_x(\mathbf{x})\} \qquad \mathit{for}\ i = 1, 2, \ldots, m \tag{48}$$

and

$$\mathbf{u_i} \in \mathit{Null}\{S_x^T(\mathbf{x})\} \qquad \mathit{for}\ i = m+1, m+2, \ldots, N_x \tag{49}$$

Proof The correlation matrix $R(\mathbf{x})$ is given by

$$R(\mathbf{x}) = S(\boldsymbol{\theta})S(\boldsymbol{\theta})^T + N_y\sigma^2 I_{N_x} \tag{50}$$

and if $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{N_x}$ are the eigenvectors corresponding to the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_{N_x}$, such that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{N_x}$

$$R(\mathbf{x})\mathbf{u} + \mathbf{i} = \lambda_i \mathbf{u_i} \qquad \mathit{for}\ i = 1, 2, \ldots, N_x \tag{51}$$

which implies that $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{N_x}$ are also the eigenvectors of $S(\boldsymbol{\theta})S(\boldsymbol{\theta})^T$, that is,

$$S(\boldsymbol{\theta})S(\boldsymbol{\theta})^T \mathbf{u_i} = \nu_i \mathbf{u_i} \qquad \mathit{for}\ i = 1, 2, \ldots, N_x \tag{52}$$

Therefore,

$$\lambda_i = \nu_i + N_y\sigma^2 \qquad \mathit{for}\ i = 1, 2, \ldots, N_x \tag{53}$$

However, $S_x(\mathbf{x})$ and $S_x(\mathbf{y})$ are of full rank m, therefore $S(\boldsymbol{\theta})S(\boldsymbol{\theta})^T$ has $N_x - m$ zero eigenvalues, that is,

$$\lambda_i = \begin{cases} \nu_i + N_y\sigma^2 & \mathit{for}\ i = 1, 2, \ldots, m \\ N_y\sigma^2 & \mathit{for}\ i = m+1, m+2, \ldots, N_x \end{cases}$$

Since the closed form of $S(\boldsymbol{\theta})S(\boldsymbol{\theta})^T$ is given by

$$S(\boldsymbol{\theta})S(\boldsymbol{\theta})^T = S_x(\mathbf{x})AS_x^T(\mathbf{y})S_x(\mathbf{y})AS_x^T(\mathbf{x}), \tag{54}$$

and $S_x(\mathbf{x})$ and $A\ S_x^T(\mathbf{y})\ S_x(\mathbf{y})\ A$ are both of full rank m, the eigenvectors corresponding to the $N_x - m$ smallest eigenvalues of $R(\mathbf{x})$ are in the null space of $S_x^T(\mathbf{x})$. The other m eigenvectors that are associated with the m largest eigenvalues are in the range space of $S_x(\mathbf{x})$.

Similar properties are enjoyed by the correlation matrix of $\mathbf{y}$, $R(\mathbf{y})$.

Theorem 2. Let the rank of $S_x(\mathbf{x})$ and $S_y(\mathbf{y})$ be m. Furthermore, let $\lambda_1, \lambda_2, \ldots, \lambda_{N_y}$ be the eigenvalues of $R(\mathbf{y})$ and $\nu_1, \nu_2, \ldots, \nu_{N_y}$ be the eigenvalues of $S(\boldsymbol{\theta})^T S(\boldsymbol{\theta})$. Then

$$\lambda_i = \begin{cases} \nu_i + N_x\sigma^2 & \text{for } i = 1, 2, \ldots, m \\ N_x\sigma^2 & \text{for } i = m+1, m+2, \ldots, N_y \end{cases} \tag{55}$$

Furthermore, if $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_{N_y}}$ are the corresponding eigenvectors of $R(\mathbf{y})$,

$$\mathbf{v_i} \in \text{Range}\ \{S_y(\mathbf{y})\} \qquad \text{for } i = 1, 2, \ldots, m \tag{56}$$

and

$$\mathbf{v_i} \in \text{Null}\ \{S_y^T(\mathbf{y})\} \qquad \text{for } i = m+1, m+2, \ldots, N_y \tag{57}$$

The eigenvalues and the eigenvectors of the correlation matrices $R(\mathbf{x})$ and $R(\mathbf{y})$ appear very naturally in the singular value decomposition (SVD) of matrix D (e.g., see Ref. 18).

Singular Value Decomposition. Let D be a $N_x \times N_y$ matrix with $Rank(D) = m$, then there exist unitary matrices U, V and a diagonal matrix Σ such that

$$U^T DV = \begin{bmatrix} \sum & 0 \\ 0 & 0 \end{bmatrix} \tag{58}$$

where

$$\sum = diag(\sigma_1, \sigma_2, \ldots, \sigma_m) \tag{59}$$

and $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_m > 0$. The numbers $\sigma_1, \sigma_2, \ldots, \sigma_m$, constituting the elements of diagonal matrix Σ, are called the *singular values* of the matrix D. The columns of the unitary matrix U are called the *left singular vectors* of D. Similarly, the columns of the unitary matrix V are called the *right singular vectors* of D. The right singular vectors are the eigenvectors of $D^T D$, and the left singular vectors are the eigenvectors of DD^T.

The unitary matrices U and V have $\mathbf{u_i}$ and $\mathbf{v_i}$ as their columns, respectively. In the noise-free case, the diagonal elements of Σ are $\sqrt{\nu_i}$, where ν_i are the m nonzero eigenvalues of $S(\boldsymbol{\theta})S(\boldsymbol{\theta})^T$.

Signal and Noise Subspaces. Let U and V be partitioned so that

$$U = [U_s | U_n] \text{ and } V = [V_s | V_n] \tag{60}$$

so that U_s and V_s contain the singular vectors corresponding to the m largest singular values. U_n and V_n, on the other hand, contain the singular vectors corresponding to the *zero* singular values. The columns of U_s and V_s are said to span the signal subspace, whereas the columns of U_n and V_n span the noise subspace. In fact, it can be shown straightforwardly that

$$D = U_s \sum V_s^T \tag{61}$$

by substituting the partitioned forms of U and V into Eq. (58). Therefore, D can be expressed in the signal subspace singular vectors only.

Algorithms

Source Number Estimation. For m sources with different x and y coordinates, the ranks of $S_x(\mathbf{x})$ and $S_y(\mathbf{y})$ will both be m. In this case, the number of sources can be estimated from either Eq. (47) or (55) by determining the number of eigenvalues of $R(\mathbf{x})$ and $R(\mathbf{y})$ that are greater than $N_y\sigma^2$ and $N_x\sigma^2$, respectively. When k of the point sources have identical x or y coordinates, one of the matrices $S_x(\mathbf{x})$ and $S_y(\mathbf{y})$ will have rank $m - k$. The number of sources can still be estimated by first determining the number of eigenvalues of $R(\mathbf{x})$ and $R(\mathbf{y})$ that are greater than $N_y\sigma^2$ and $N_x\sigma^2$, respectively. Then the larger of the two results is declared as the estimate of the number of sources. Similar statements hold for the case in which k_x of the x coordinates and k_y of the y coordinates are the same.

Noise Subspace Algorithm. Equation (49) implies that the space spanned by the columns of U_n is orthogonal to the space spanned by the columns of $S_x(\mathbf{x})$, that is,

$$S_x^T(\mathbf{x})U_n = 0 \tag{62}$$

Similarly, Eq. (57) implies that

$$S_y^T(\mathbf{y})V_n = 0 \tag{63}$$

However, because of the presence of noise in the eigenvector estimates that span the noise and the signal subspaces, the orthogonality conditions above will not in general hold. One will have to find the parameter vectors $\mathbf{x}$ and $\mathbf{y}$ that most closely approximate the orthogonality conditions given by Eqs. (62) and (63). These will be given as the spectral peaks of the functions $\rho(\mathbf{x})$ and $\rho(\mathbf{y})$ such that

$$\rho(\mathbf{x}) = \frac{1}{\mathbf{s}_x^T(\mathbf{x})U_n U_n^T \mathbf{s}_x(\mathbf{x})} \tag{64}$$

and

$$\rho(\mathbf{y}) = \frac{1}{\mathbf{s}_y^T(\mathbf{y})V_n V_n^T \mathbf{s}_y(\mathbf{y})} \tag{65}$$

This algorithm is a *noise subspace* algorithm, since it involves the property of the noise subspace.

Signal Subspace Algorithm. On the other hand, Eq. (48) implies that the columns of U_s can be written as a linear combination of columns of $S_x(\mathbf{x})$, that is,

$$U_s = S_x(\mathbf{x})H_1 \tag{66}$$

where H_1 is a $m \times m$ unknown coefficient matrix. A similar expression can be written for V_s so that

$$V_s = S_y(\mathbf{y})H_2 \tag{67}$$

The parameter vectors $\mathbf{x}$, $\mathbf{y}$ and the coefficient matrices H_1 and H_2 should be chosen so that Eqs. (66) and (67) are satisfied. Therefore, one can at most search for the parameter vectors that will minimize a chosen norm of the error matrices $U_s - S_x^T(\mathbf{x})H_1$ and $V_s - S_y^T(\mathbf{y})H_2$. One such widely used norm is the Frobenius norm of the error matrices

$$\rho(\mathbf{x}) = \|U_s - S_x(\mathbf{x})H_1\|_2^2 \tag{68}$$

and

$$\rho(\mathbf{y} = \|V_s - S_y(\mathbf{y})H_2\|_2^2 \tag{69}$$

where Eq. (68) is to be minimized with respect to $\mathbf{x}$ and Eq. (69) is to be minimized with respect to $\mathbf{y}$. Unfortunately, a closed form solution to the minimization problems above almost never exists because of the nonlinear manner the matrices $S_x(\mathbf{x})$ and $S_y(\mathbf{y})$ depend on the unknown parameter vectors $\mathbf{x}$ and $\mathbf{y}$. This being the case, a nonlinear programming method must be used. For this problem, we also used the Gauss–Newton method with QR decomposition. The Gauss–Newton method, as a descent method, expects "good" initial estimates for the parameter vectors to be estimated. The initial estimates are supplied by the Sequential Orthogonal Projection method.

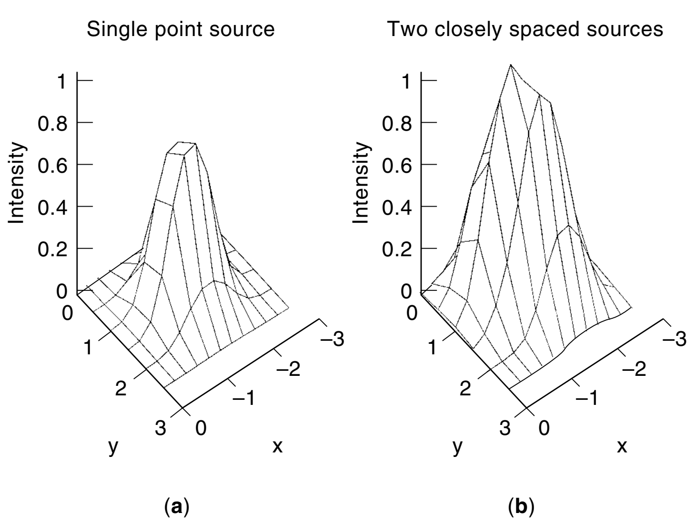

Figure 2. The intensity function generated on the focal plane: (a) single source at (−1.40Δ, 1.30Δ); (b) two closely spaced sources at (−1.070Δ, 0.887Δ) and (−1.671Δ, 1.488Δ). From (12).

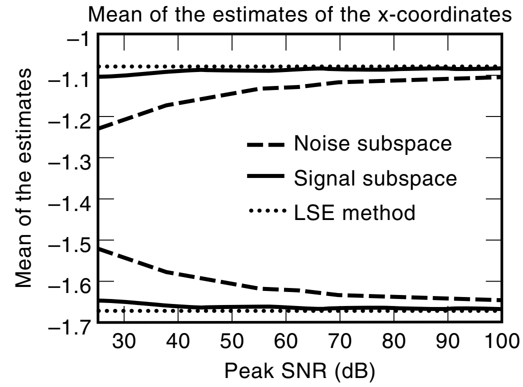

Figure 3. The mean of the estimates of the x-coordinates. From (12).

SPECIAL CASE: GAUSSIAN POINT SOURCES

In this section, we will test the effectiveness of the proposed multiple source location and outlier detection algorithms for a specific application. We assume that the point spread function of the projected focal plane IR intensity density function of a point source located at (x_k, y_k, z_k) is specified by the commonly employed symmetric Gaussian function

$$i(x, y) = \frac{1}{2\pi\omega^2} e^{-[(x-x_k)^2+(y-y_k)^2]/2\omega^2} \tag{70}$$

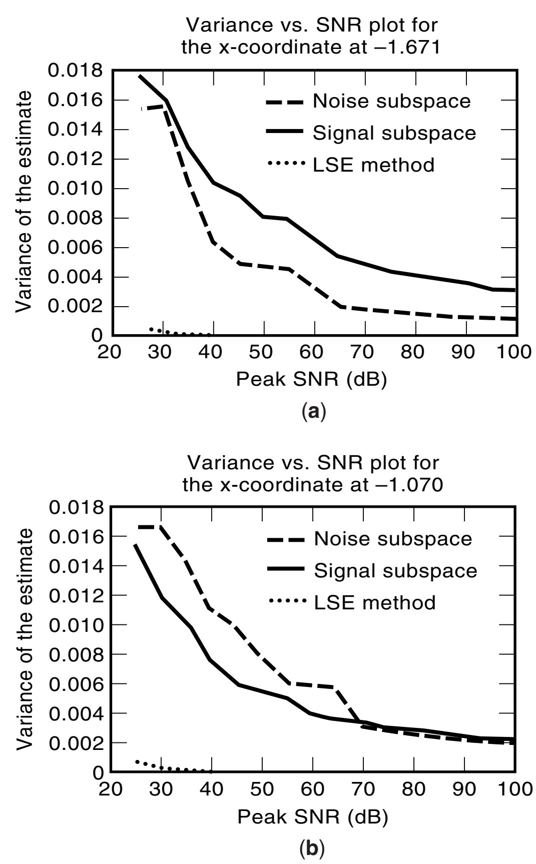

Figure 4. The variances of the estimates for different estimators under Gaussian noise only: (a) x-coordinate at −1.671; (b) x-coordinate at −1.070. From (12).

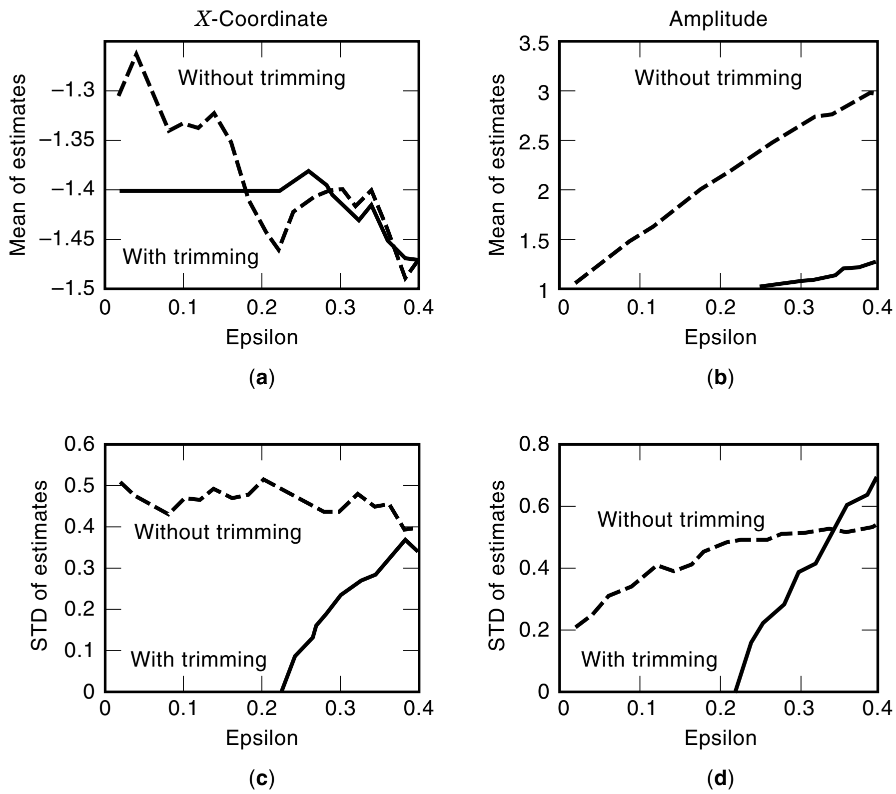

Figure 5. The LSE estimates with and without outlier detection (trimming) algorithm for different contamination levels. Triangular noise only. From (12).

where the blur width parameter, ω, is assumed to be known and controls the spread of the function. This point source illuminates an array of square-shaped IR detector cells with sides of length Δ that are parallel to the x and y axes. The response of such a detector cell with its center located at (x_c, y_c) to the IR intensity function [Eq. (70)] is given by

$$s(x_c, y_c) = \frac{1}{2\pi\omega^2} \int_{y_c-0.5*\Delta}^{y_c+0.5*\Delta} \int_{x_c-0.5*\Delta}^{x_c+0.5*\Delta} e^{-(x-x_k)^2+(y-y_k)^2/2\omega^2}\, dx\, dy \qquad (71)$$

It is clear that this detector cell response is equal to the volume of the two-dimensional point spread function [Eq. (70)] above the square-shaped detector cell surface. Although a closed form solution for this integral does not exist, it is possible to represent this integral in terms of the Gaussian cumulative distribution function whose values are available in numerical tables. Thus, we have

$$\begin{aligned} s(x_c, y_c) = {} & \left[\Phi\left(\frac{x_1 - x_c + 0.5\Delta}{\omega}\right) - \Phi\left(\frac{x_1 - x_c - 0.5\Delta}{\omega}\right)\right] \\ & \left[\Phi\left(\frac{y_1 - y_c + 0.5\Delta}{\omega}\right) - \Phi\left(\frac{y_1 - y_c - 0.5\Delta}{\omega}\right)\right] \end{aligned} \qquad (72)$$

where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt \qquad (73)$$

Examination of Eq. (72) indicates that the detector cell response is equal to the product of a function dependent on x coordinates with a function dependent on y coordinates. Thus $s(x_c, y_c)$ is a *separable* function of x and y. The plot of this two-dimensional "Gaussian-like" function corresponding to a unit amplitude point source located at $(-1.400\Delta, 1.300\Delta)$ is shown in Fig. 2(a).

In the simulations, the following staggered detector cell array configuration consisting of eighteen square detectors of size Δ is used:

$$\begin{array}{cccccc} (0,0) & & (0,\Delta) & & (0,2\Delta) & \\ & (\frac{\Delta}{2}, \frac{\Delta}{2}) & & (\frac{\Delta}{2}, \frac{3\Delta}{2}) & & (\frac{\Delta}{2}, \frac{5\Delta}{2}) \\ (\Delta,0) & & (\Delta,\Delta) & & (\Delta,2\Delta) & \\ & (\frac{3\Delta}{2}, \frac{\Delta}{2}) & & (\frac{3\Delta}{2}, \frac{3\Delta}{2}) & & (\frac{3\Delta}{2}, \frac{5\Delta}{2}) \\ (2\Delta,0) & & (2\Delta,\Delta) & & (2\Delta,2\Delta) & \\ & (\frac{5\Delta}{2}, \frac{\Delta}{2}) & & (\frac{5\Delta}{2}, \frac{3\Delta}{2}) & & (\frac{5\Delta}{2}, \frac{5\Delta}{2}) \end{array}$$

Assume that the main lobe of a static point source illuminating the focal plane remains in a given cell for four samples while the plane is moved along the negative x-axis with constant velocity. With three detectors in a row, twelve samples are obtained in the x-direction. Since every other column is shifted by a half detector width and there are three detectors in each column, the data can be expressed by a matrix with 12 rows and 6 columns.

We assume there are two unit amplitude stationary point sources located at $(-1.0700\Delta, 0.8870\Delta)$ and $(-1.6710\Delta, 1.4880\Delta)$ on the focal plane. With this choice of coordinates, the two point sources are separated by a distance of 0.85Δ from each other. The size of a square detector, Δ, is chosen so that a unit point source located at the center of a cell will induce a response of 0.86 on the detector. The noise free data matrix obtained by this configuration is shown in Fig. 2(b).

The data matrix is then corrupted by additive white Gaussian noise at different maximum signal-to-noise-ratio

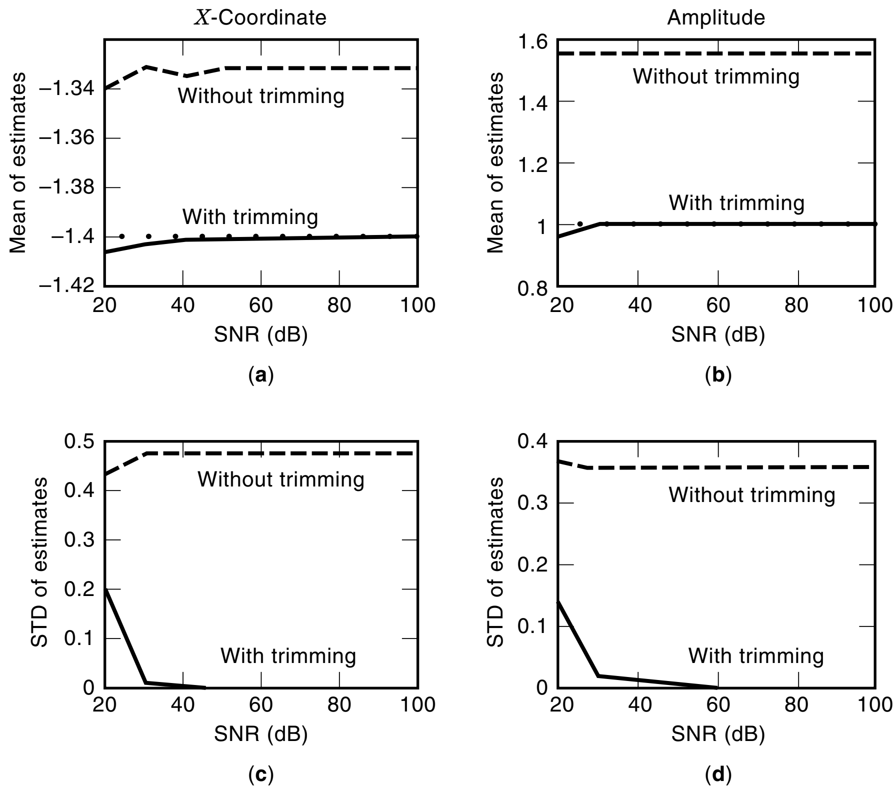

Figure 6. The outlier detection (trimming) algorithm with mixture noise. Different Gaussian noise levels (SNRs)/10% contamination with triangular noise. From (12).

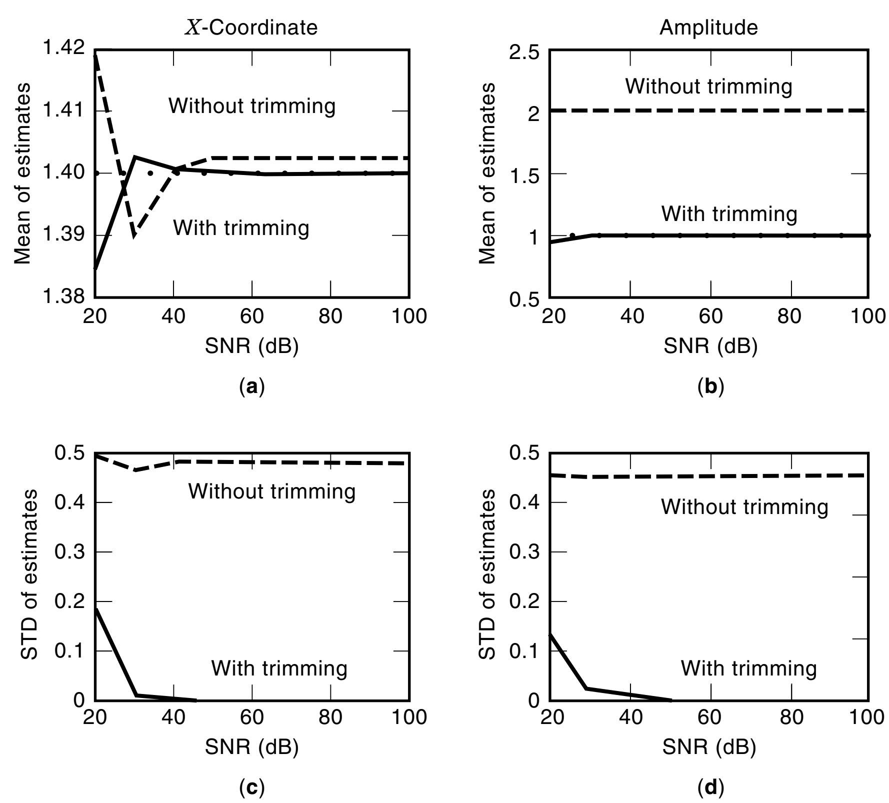

Figure 7. The outlier detection (trimming) algorithm with mixture noise. Different Gaussian noise levels (SNRs)/18% contamination.

(SNR) levels. The relationship between the SNR and the standard deviation of the Gaussian noise is given by the expression

$$\mathrm{SNR}_{\mathrm{peak}} = 10 \log \frac{0.86}{\sigma^2} \tag{74}$$

At various signal-to-noise ratio levels, 100 trial runs of the experiment are performed and the estimates of the x and y coordinates obtained by each method are recorded.

As a next step, the effectiveness of the outlier detection scheme is tested with a unit amplitude point source located at (-1.400Δ, 1.300Δ). The two-dimensional data matrix for this point source is depicted in Fig. 2(a). Samples of the contaminating distribution are generated from a triangular distribution with mean two and variance one. With this modeling, the values noise samples take are one to three times the amplitude of the point source. One hundred trial runs of the experiment are performed at different contamination levels ϵ. The estimates of the location coordinates and amplitudes obtained with and without the outlier detection scheme are recorded. The algorithm is also tested with a mixture noise. In this case, the noise samples are drawn from a Gaussian distribution with probability $(1 - \epsilon)$, and from a Gaussian plus triangular distribution with probability ϵ. One hundred trial runs of the experiment are performed at various signal-to-noise ratios and at three levels of contamination ($\epsilon = 0.02$, 0.10, 0.18) of triangular noise.

Simulation Results

The performance of two subspace methods and the LSE estimation with a Gram-Schmidt orthogonalization step are tested. The signal and the noise subspace algorithms employ the separability property of the Gaussian-like intensity function. The LSE technique is also applicable to inseparable signals. The arithmetic means of the estimates of the 100 trials performed at each signal-to-noise-ratio are given in Fig. 3. The correct values of the location parameters are -1.070Δ and -1.671Δ. Among the subspace methods, the signal subspace algorithm outperformed the noise subspace algorithm. However, the estimates obtained through the method described here were virtually unbiased and outperformed both of the subspace methods. When the variances of the estimates are compared, the two subspace methods performed similarly, whereas the algorithm with Gram-Schmidt orthogonalization converged with least variance at all signal-to-noise-ratios and is highly recommendable in Gaussian noise environments. The variance of the estimates of the x coordinates at different signal-to-noise ratios are depicted in Fig. 4(a) and (b). The Cramer–Rao lower bounds (CRLB) for the estimates are calculated by applying the general results of Stoica and Nehorai (17) to the Gaussian point source location problem. The mean square error of the estimates obtained by the minimization of the squared error achieved the CRLB at high signal-to-noise ratios. Similar results are obtained for the y coordinates and the amplitudes. In Ref. 18, simulation results for the means square errors and the CRLB are given for varying source separations; the authors observe that the LSE method approaches the CRLB as the point source separation increases.

In the presence of nonzero mean triangular distribution only, the estimates obtained through the minimization of the square error are severely degraded. As depicted in Fig. 5, the estimates of location and amplitude without trimming are not only biased, but have high variance as well. When the trimming algorithm is employed, the means of estimates of the x location and amplitude parameters of the point source are equal to their actual values of -1.400Δ and 1.000 for contamination levels up to 20%. Moreover, the corresponding standard deviations of the estimates are zero, indicating perfect retrieval for one hundred trials at every contamination level. Figure 6 depicts the statistical properties of the LSE estimates with and without trimming for a mixture noise. In this particular case, 90% of the noise samples are from a Gaussian distribution at the given signal-to-noise ratio, and the remaining 10% are samples from the triangular distribution in addition to Gaussian noise. Another experiment for the case in which 18% of the noise samples are from the triangular distribution is performed and the results are presented in Fig. 7. For both cases, the trimming algorithm is very effective when the Gaussian portion of the noise samples are at maximum signal-to-noise ratios higher than 30 dB.

BIBLIOGRAPHY

1. D. A. Scribner, M. R. Kruer, and J. M. Killany, Infrared focal plane array technology, *Proc. IEEE,* **79**: 68–85, 1991.
2. R. A. Ballingall, Review of infrared focal plane arrays, *Proc. SPIE 1320:* 70–87, 1990.
3. P. W. Kruse, Uncooled IR focal plane arrays, *Proc. SPIE 2552:* 556–563, 1995.
4. I. M. Baker et al., Infrared detectors for the year 2000, *Infrared Review,* **9** (2): 50–60, 1996.
5. G. H. Golub and V. Pereyra, The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate, *SIAM J. Numerical Analysis,* 413–432, April 1973.
6. J. E. Dennis, Jr. and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations,* Englewood Cliffs, NJ: Prentice Hall, 1983.
7. D. G. Luenberger, *Introduction to Linear and Nonlinear Programming,* Reading, MA: Addison-Wesley, 1965.
8. J. A. Cadzow, Least squares error modeling with signal processing applications, *IEEE Trans. Acoust. Speech Signal Process. Magazine,* 12–31, October 1990.
9. Y. Yardımcı and J. A. Cadzow, Multiple point source location via IR array processing methods, *Proc. 22nd Conf. System Theory,* 230–234, 1990.
10. Y. Yardımcı and J. A. Cadzow, Direction-of-arrival estimation in the presence of sensor model errors, *J. Franklin Institute,* August 1995.
11. Y. Yardımcı, J. A. Cadzow, and A. E. Çetin, Robust signal modeling through nonlinear least squares, *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.* 1994.
12. Y. Yardımcı and J. A. Cadzow, High-resolution algorithms for locating closely spaced objects via infrared focal-plane arrays, *Optical Engineering,* **33** (10): 3315–3323, 1994.
13. I. Ziskind and M. Wax, Maximum likelihood localization of multiple sources by alternating projection, *IEEE Trans. Acoust. Speech Signal Process,* **ASSP-36**: 1553–1560, 1988.
14. V. Barnett and T. Lewis, *Outliers in Statistical Data,* Chichester: Wiley, 1984.
15. V. F. Pisarenko, The retrieval of harmonics from a covariance function, *Geophys. J. Roy. Astron. Soc.,* **33**: 347–366, 1973.
16. R. Schmidt, Multiple emitter location and signal parameter estimation, *Proc. RADC Spectral Estimation Workshop,* 243–256, 1979.

17. P. Stoica and A. Nehorai, MUSIC, maximum likelihood, and Cramer-Rao bound, *IEEE Trans. Acoust. Speech Signal Process.* **ASSP-37**: 720–740, 1989.

18. J. T. Reagan and T. J. Abatzoglou, Model-based superresolution CSO processing. In O. E. Drummond (ed.), *Proc. of SPIE 1954,* 204–218, 1993.

YASEMIN YARDıMCı
Vanderbilt University
JAMES A. CADZOW
Vanderbilt University

FOLDED MONOPOLE ANTENNAS. See MONOPOLE ANTENNAS.

FORCE MEASUREMENT. See WEIGHING.

FORCE SENSORS. See DYNAMOMETERS.

Another principle of forecasting is that forecasts are more accurate for groups or families of items rather than for individual items themselves. Because of pooling of variances, the behavior of group data can have very stable characteristics even when individual items in the group exhibit high degrees of randomness. Consequently, it is easier to obtain a high degree of accuracy when forecasting groups of items rather than individual items themselves.

Finally, forecasts are more accurate for shorter than longer time horizons. The shorter the time horizon of the forecast, the lower the uncertainty of the future. There is a certain amount of inertia inherent in the data, and dramatic pattern changes typically do not occur over the short run. As the time horizon increases, however, there is a much greater likelihood that a change in established patterns and relationships will occur. Therefore, forecasters cannot expect to have the same degree of forecast accuracy for long range forecasts as they do for shorter ranges.