# Statistical Performance Analysis of Super-Resolution

Dirk Robinson and Peyman Milanfar, *Senior Member, IEEE*

***Abstract*—Recently, there has been a great deal of work developing super-resolution algorithms for combining a set of low-quality images to produce a set of higher quality images. Either explicitly or implicitly, such algorithms must perform the joint task of registering and fusing the low-quality image data. While many such algorithms have been proposed, very little work has addressed the performance bounds for such problems. In this paper, we analyze the performance limits from statistical first principles using Cramér–Rao inequalities. Such analysis offers insight into the fundamental super-resolution performance bottlenecks as they relate to the subproblems of image registration, reconstruction, and image restoration.**

***Index Terms*—Cramer–Rao (CR) bounds, Fisher information, image reconstruction, image restoration, performance limits, super-resolution.**

## I. Introduction

THE last decade has seen a great deal of work in the development of algorithms addressing the problem of super-resolution. We refer the interested reader to [1] for a broad review of the work in this area. In general, the problem of super-resolution can be expressed as that of combining a set of aliased, noisy, low-resolution, blurry images to produce a higher resolution image or image sequence. With some simplifying assumptions, the estimation problem is typically divided into the tasks of first registering the low-resolution images with respect to the coordinate system of the desired high-resolution image, followed by a reconstruction or fusion of the low-resolution data combined with deblurring and interpolation, to produce the final high-resolution image. Historically, most researchers in super-resolution have tended to focus on the latter reconstruction/restoration stages assuming that generic image registration algorithms could be trusted to produce estimates with a high level of accuracy. Relatively recently, researchers have noted the importance of solving the estimation problems of image registration and super-resolution in a joint fashion [2]–[4]. In this paper, we study this relationship between the task of image registration and image reconstruction.

Specifically, we analyze the joint problem of image registration and super-resolution (high-resolution image reconstruction/restoration) in the context of fundamental statistical performance limits. Little work has addressed performance limits for the problem of super-resolution. Relevant works are [5] and [6]. Both works study the problem of super-resolution from an algebraic perspective reducing all super-resolution algorithms to that of solving large systems of linear equations. Essentially, [5] studies the numerical conditioning of the linear system of equations characterizing the forward measurement model. In [5], it is assumed that the image registration parameters are known perfectly. In some sense, the work has little to do with the performance limits of super-resolution and more to do with quantifying the ill posedness of image deconvolution or deblurring. That work relates the tendency of the system of linear equations to become more ill posed as the downsampling factor increases. In this paper, we show how such analysis makes implicit assumptions which lead to an incomplete understanding of the super-resolution problem. Similar assumptions are made in [6] to derive lower bounds on resolution enhancement performance. In [6], performance limits are derived based on inequalities for matrix perturbations. While some of the matrix perturbation analysis incorporates the notion of uncertainty associated with image registration, the entire analysis continues to assume that image registration is an entirely independent step in the super-resolution process. One observation noted in [6] is that, for most imaging applications, the enhancement factor of 1.6 is "unbreakable." In this paper, we explore the conditions under which resolution enhancement factors as great as 5 are achieved by some super-resolution algorithms [1].

In this paper, we study the performance limits for super-resolution from a statistical perspective, enabling us to bound estimator performance in terms of mean-square error (MSE) using Cramér–Rao (CR) inequalities. To do so, we frame the super-resolution problem as a parametric estimation problem where MSE becomes a useful measure of performance. Furthermore, we study the performance limits for super-resolution in its entirety, which includes analysis of both the image registration and reconstruction/restoration problems. We note that the CR inequality has been used recently to study the performance limits associated with the related problem of motion-free superresolution [7], whereby a higher resolution image is reconstructed from multiple frames, each having a different blurring point-spread function (PSF).

Using CR inequalities, we explore super-resolution performance limits as they relate to various imaging system parameters such as the downsampling factor, signal-to-noise ratio (SNR), and the PSF. In addition, our analysis exposes the signal-dependent aspect of performance, such as the class of images under observation, as well as the set of motion vectors observed in the low-resolution data. We shed light on important application-level questions regarding the necessary number

Manuscript received December 7, 2004; revised May 21, 2005. This work was supported in part by AFOSR Grant F49620-03-1-0387. D. Robinson performed this work while with the Department of Electrical Engineering, University of California at Santa Cruz. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Mario A. T. (G. E.) Figueiredo.

D. Robinson is with Ricoh Innovations, Menlo Park, CA 94025 USA (e-mail: dirkr@ee.ucsc.edu).

P. Milanfar is with the Department of Electrical Engineering, Baskin School of Engineering, University of California at Santa Cruz, Santa Cruz, CA 95064 USA (e-mail: milanfar@ee.ucsc.edu).

Digital Object Identifier 10.1109/TIP.2006.871079

of measurement frames and prior information necessary to achieve a given level of performance.

This paper is organized as follows. In Section II, we present a parametric model for super-resolution and introduce the CR inequality with which we can bound the MSE for large classes of estimators. In Section III, we derive the information matrices for super-resolution as they relate to the observed data and prior information models. We reformulate the problem in the Fourier domain as a means to efficiently compute the CR bounds. In Section IV, we analyze the performance bounds for a variety of scenarios offering insight into the inherent challenges and tradeoffs in super-resolution. Finally, in Section V, we summarize the contributions of this paper and suggest future directions.

## II. CR Bounds for Super-Resolution

For the general super-resolution problem, it is assumed that one is given a set of low-resolution images which consist of noisy, warped, blurred, and downsampled versions of an unknown high-resolution image. The simplest, and perhaps most commonly utilized, warping model assumes that the motion between frames is captured by a global shift or a translation [1]. To maintain focus and gain intuition, we concentrate on the case of translational motion. For such an assumption, we represent the forward process by a linear function of the unknown image as

$$\mathbf{y}_k = \mathbf{DHF}(\mathbf{v}_k)\mathbf{x} + \mathbf{e}_k. \tag{1}$$

The vectors $\mathbf{y}_k$ represent the samples of the $k$th measured image scanned in some fashion to form $N_L$-dimensional vectors. Likewise, $\mathbf{x}$ represents the unknown desired high-resolution image similarly scanned to form a $N_H$-dimensional vector. The matrix $\mathbf{D}$ captures the downsampling operation, $\mathbf{H}$ the blurring operation due to the imaging system PSF, and $\mathbf{F}(\mathbf{v}_k)$ the translational motion operation with $\mathbf{v}_k = [v_{k_1}, v_{k_2}]^T$ being the unknown translation parameters for a particular frame. Finally, $\mathbf{e}_k$ represents the vector of additive white Gaussian measurement noise with variance $\sigma^2$.

For the purpose of this paper, we make several assumptions about the forward process. First, we assume the unknown high-resolution image $\mathbf{x}$ to be reconstructed is a periodic bandlimited image sampled above the Nyquist rate. From this assumption, the matrix $\mathbf{F}(\mathbf{v}_k)$ (which we will refer to as $\mathbf{F}_k$) representing the translational shift of the image $\mathbf{x}$, reflects a convolution operation with a shifted two-dimensional (2-D) sinc function. Such a motion formulation allows arbitrary, possibly noninteger shifts. The matrix $\mathbf{F}_k$ has the property that $\mathbf{F}_k^T\mathbf{F}_k = \mathbf{I}$ where $\mathbf{I}$ is the identity matrix. In other words, shifting the image followed by a shift in the reverse direction does not change the pixel values of the high-resolution image. Furthermore, we note that when the motion vector $\mathbf{v}_k$ reflects integer shifts, then the matrix $\mathbf{F}_k$ is simply a permutation of the identity matrix $\mathbf{I}$. Second, we assume that the blurring operation, and, hence, the imaging system's PSF is symmetric and spatially invariant and can be represented by a convolution. From this assumption, we can represent the blurring operator $\mathbf{H}$ with a block circulant matrix.[1] Third, we assume that the downsampling operation is based on a known downsampling factor $M$ where $(N_L/N_H) = (1/M)$. For our purposes, we assume that the downsampling factor $M$ is an integer. Thus, $\mathbf{D}$ is an $N_L$ by $N_H$ matrix representing the downsampling operation. Finally, in our formulation, we suppose that $K+1$ low-resolution measured images are available. Without loss of generality, we assume that the initial image $\mathbf{y}_0$ dictates the coordinate system so that $\mathbf{F}_0 = \mathbf{I}$, and, hence, we only have to estimate $K$ unknown translation vectors $\mathbf{v}_k$ during the super-resolution process for a given set of $K+1$ low-resolution frames.

In what follows, we characterize the fundamental performance limits for super-resolution in terms of CR bounds. Essentially, the CR bounds characterize, from an information theoretic standpoint, the "difficulty" with which a set of parameters can be estimated by examining the given model [8]. In general, the CR inequalities provide lower bounds on the MSE for a large class of estimators $\hat{\theta}$ of an unknown parameter vector $\theta$ from a given set of data represented by the vector $\Phi$. The CR inequalities used in this paper are of the form

$$E[(\hat{\theta}-\theta)(\hat{\theta}-\theta)^T] \geq \mathbf{J}^{-1} \tag{2}$$

where the left-hand side error correlation matrix is called the MSE matrix since the diagonal terms of $E[(\hat{\theta}-\theta)(\hat{\theta}-\theta)^T]$ represent the MSE for the individual parameter components. The right-hand side is referred to as the CR bound and the matrix $\mathbf{J}$ is called the information matrix. The inequality indicates that the difference between the MSE (left side) and the CR bound (right side) will be a positive semidefinite matrix. In this paper, we focus on two estimation regimes. In the first, we study the performance limits of conditional estimation for a particular $\theta$ under the assumption that no prior information is available about $\theta$. In the second, we study the performance limits when $\theta$ is an unknown parameter with a known prior distribution.

In the case of conditional estimation, the CR bound limits the performance for *any* unbiased estimator. For instance, the class of maximum likelihood (ML) estimators fall into this category. In this regime, the CR bound limits performance in a conditional MSE sense. The conditional CR bound is given by

$$E_{\Phi\,|\,\theta}[(\hat{\theta}-\theta)(\hat{\theta}-\theta)^T] \geq \mathbf{J}^{-1}. \tag{3}$$

The information matrix in this case is called the Fisher information matrix (FIM) and is defined as

$$\{\mathbf{J}\}_{i,j} = -E_{\Phi\,|\,\theta}\left[\frac{\partial^2 l(\theta\,|\,\Phi)}{\partial\theta_i\partial\theta_j}\right] \tag{4}$$

where $l(\theta\,|\,\Phi)$ is the log-likelihood of the measured data $\Phi$ for a given value of the unknown parameter $\theta$. In this case, the Fisher information matrix is a function of the unknown parameter $\theta$. As we shall see, the bound in this form offers us intuition on the performance in estimating a particular signal.

In the second regime, we use a Bayesian CR bound to study the performance in estimating an unknown parameter coming from a known distribution. In this case, the inequality (3) bounds the MSE over the joint distribution of the parameter and noise

$$E_{\Phi,\theta}[(\hat{\theta}-\theta)(\hat{\theta}-\theta)^T] \geq \mathbf{J}^{-1}. \tag{5}$$

[1]We leave the generalization to more complicated and possibly unknown PSFs for future research.

For a given log-prior distribution on the unknown parameter $l(\theta)$, the information matrix in the Bayesian regime is given by

$$\begin{aligned}\{\mathbf{J}\}_{i,j} &= -E_{\mathbf{\Phi},\theta}\left[\frac{\partial^2 l(\theta \mid \Phi)}{\partial\theta_i \partial\theta_j}\right] - E_{\theta}\left[\frac{\partial^2 l(\theta)}{\partial\theta_i \partial\theta_j}\right] \\ &= \{\mathbf{J}_d\}_{i,j} + \{\mathbf{J}_p\}_{i,j}. \end{aligned} \tag{6}$$

We use the subscript $d$ to denote information arising from the measured *data* and $p$ to denote the information from the *prior* [9]. We see that in the Bayesian case, the data information matrix is the expected value of the Fisher information matrix over parameter prior distribution. In this case, the total information matrix offers a bound on the MSE over the class of estimators which are weakly biased as defined in [9], [10]. Maximum *a posteriori* (MAP) estimators are a natural class of estimators falling into this category [9].

In certain operating scenarios, the FIM of (4) can be very poorly conditioned or even rank deficient. In such situations, prior knowledge about the unknown parameter is essential to solving the problem. Essentially, it is this prior knowledge, learned from a set of training instances, that the authors of [5] use to "break" the limits of super-resolution. In fact, the CR bound offers a convenient mechanism for understanding the utility of such prior knowledge while still offering a fundamental performance limit. In other words, prior information does not allow one to break the performance limits but, instead, makes the fundamental limit more favorable. Indeed, an interesting related question, which we do not treat here, is "which $l(\theta)$ makes the overall performance limits most favorable?" That is, what is the most informative prior for a given observation model.

Focusing the discussion now on the super-resolution estimation problem, we have two sets of unknown parameters in the model; namely, the unknown high-resolution image $\mathbf{x}$ and the set of unknown motion vectors $\{\mathbf{v}_k\}$. To simplify the notation, we will represent the entire set of unknown motion vectors as $\underline{v} = [v_{1_1}, v_{1_2}, \ldots, v_{K_1}, v_{K_2},]^T$. As such, the unknown parameter vector $\theta$ is given by $\theta = [\mathbf{x}^T, \underline{v}^T]^T$. Depending on the application, it is natural to distinguish the problem of estimating the translational motion parameters $\underline{v}$ from the estimation of the image signal $\mathbf{x}$. For instance, in super-resolution, the motion parameters $\underline{v}$ represent what are known as unknown *nuisance* parameters [9]. In other words, estimates of the motion parameters are necessary to reconstruct the image $\mathbf{x}$ but otherwise have no intrinsic importance. Because of this dichotomy, we analyze our information matrix $\mathbf{J}$ using the following partitioned structure:

$$\mathbf{J} = \begin{pmatrix} \mathbf{J}_{xx} & \mathbf{J}_{xv}^T \\ \mathbf{J}_{xv} & \mathbf{J}_{vv} \end{pmatrix} \tag{7}$$

where the matrix $\mathbf{J}_{xx}$ captures the available information solely pertaining to the unknown image $\mathbf{x}$, $\mathbf{J}_{vv}$ the information about the motion parameters $\underline{v}$, and $\mathbf{J}_{xv}$ reflects the *information intercorrelation* between the two sets of unknown parameters. Were $\mathbf{J}_{xv} = 0$, then the problem of image reconstruction (estimation of $\mathbf{x}$) could be decoupled from the problem of image registration without any loss in performance. Since this is, in general, not the case, we argue that the problems of registration and image reconstruction must be solved in a joint fashion.[2]

Although the estimation must be performed jointly, using the block decomposition of (7), we can separate our performance analysis for the two estimation problems using the block matrix inversion principle [11]. Using this principle, the CR bound is given by

$$\mathbf{J}^{-1} = \begin{pmatrix} \mathbf{S}_{\mathbf{x}}^{-1} & \mathbf{J}_{xx}^{-1}\mathbf{J}_{xv}\mathbf{S}_{\underline{v}}^{-1} \\ \mathbf{S}_{\underline{v}}^{-1}\mathbf{J}_{xv}^T\mathbf{J}_{xx}^{-1} & \mathbf{S}_{\underline{v}}^{-1} \end{pmatrix} \tag{8}$$

where the $\mathbf{S}$ matrices are the Schur information complements are

$$\mathbf{S}_{\underline{v}} = \mathbf{J}_{vv} - \mathbf{J}_{xv}^T\mathbf{J}_{xx}^{-1}\mathbf{J}_{xv} \tag{9}$$

$$\mathbf{S}_{\mathbf{x}} = \mathbf{J}_{xx} - \mathbf{J}_{xv}\mathbf{J}_{vv}^{-1}\mathbf{J}_{xv}^T. \tag{10}$$

In this block-partitioned formulation, we see a certain symmetry of the two estimation problems. At first glance, we observe that there is a net loss in information due to the interdependence of the two estimation problems because the second terms in the right-hand sides of (9) and (10) are positive semidefinite matrices. The information matrices capture the relationship between small perturbations of the unknown signal parameters and the probability distribution of the measured data. As such, the net loss of information reflects the ambiguity arising from a small perturbation in either sets of signal parameters producing the same perturbation in the distribution. Simply put, such a structure captures the (in)ability to distinguish variations in the measured data as depending on one parameter set versus the other.[3]

Typically, for the problem of super-resolution, we are interested in a scalar performance measure reflecting the goal of restoring an entire image $\mathbf{x}$. A natural global measure of performance is given by

$$\overline{\mathrm{rmse}}(\mathbf{x}) = \left[\frac{1}{N_H}\sum_i E\left[(x_i - \hat{x}_i)^2\right]\right]^{\frac{1}{2}} \tag{11}$$

which shows the root MSE performance over the entire image $\mathbf{x}$ in units of gray levels. The corresponding scalar CR inequality over the image $\overline{\mathrm{rmse}}(\mathbf{x}) \geq T(\mathbf{x})$ is defined using the trace of the Schur information complement

$$T(\mathbf{x}) = \left(\frac{1}{N_H}\mathrm{tr}\left(\mathbf{S}_{\mathbf{x}}^{-1}\right)\right)^{\frac{1}{2}}. \tag{12}$$

This performance measure is in terms of the root average MSE per pixel (in units of gray levels). Such a performance bound has been justified and used in the past [13].

One important property of the CR bounds we will exploit momentarily is the bounds' invariance with respect to a linear transformation of the unknown parameters. Using the property that the circulant matrices $\mathbf{H}$ and $\mathbf{F}_k$ commute, a natural linear

[2]An interesting question worth noting is how, by appropriate transformation of the problem, we might block diagonalize the FIM. While this would yield a decoupling of the problem, the resulting parameters to be estimated would not longer correspond to simply the high-resolution image pixels and motion vectors.

[3]The loss in information can be explained from a geometric point of view as the orthogonal projection of the information content about one set of unknown parameters onto the linear subspace encompassing the information about the other set of parameters [12].

transformation of the unknown image $\mathbf{x}$ is given by $\mathbf{z} = \mathbf{H}\mathbf{x}$ which maps the unknown high-resolution image into a blurry version of the high-resolution image $\mathbf{z}$. The modified signal model becomes

$$\mathbf{y}_k = \mathbf{D}\mathbf{F}(\mathbf{v}_k)\mathbf{z} + \mathbf{e}_k. \tag{13}$$

Such a transformation is commonly used in developing fast sequential super-resolution algorithms [1], [14]. In both works, the transformation was used to break the problem of super-resolution into the two subproblems of image reconstruction or fusion (estimating $\mathbf{z}$ from the data $\{\mathbf{y}_k\}$) and image restoration (estimating $\mathbf{x}$ from $\hat{\mathbf{z}}$). Such an approach is justified by the invariance property of the ML and MAP estimators for linear transformations [8]. Because this transformation is linear, it can be shown that the Schur information complements for $\mathbf{x}$ and $\mathbf{z}$ are related according to

$$\mathbf{S}_x = \mathbf{H}\mathbf{S}_z\mathbf{H}. \tag{14}$$

The linearity of the transformation permits us to derive the CR bound in terms of the unknown parameters $\mathbf{z}$, essentially ignoring the PSF for the time being. This distinction is important because it allows us to distinguish between the performance limits associated with the subproblems of image registration and reconstruction with those of image restoration. In [5] and [6], this distinction is ignored by inextricably linking the imaging system's PSF with the downsampling factor $M$. For much of this paper, we focus on the joint task of estimating $\mathbf{z}$ and $\underline{\mathrm{v}}$ since the final step of image restoration can be studied somewhat independently. In what follows, we use $T(\mathbf{z}) = ((1/N_H)\mathrm{tr}(\mathbf{S}_{\mathbf{z}}^{-1}))^{(1/2)}$ to represent the scalar CR bound in estimating the blurred image $\mathbf{z}$.

Finally, to address the utility of the CR bound in studying general estimation problems, we note that the overall usefulness of a performance limit depends on its ability not only to limit, to but predict, actual estimator performance. For example, we might trivially bound MSE performance as $\mathrm{MSE}(\theta) \geq \mathbf{0}$. While such a bound is provably correct, it offers no useful information about the estimation problem. The CR bounds, however, are known to be asymptotically attainable under certain conditions [9]. While there is no guarantee that such estimators are realizable, their existence does offer hope for predicting performance for a wide class of estimators.

## III. Information Matrices for Super-Resolution

In this section, we derive the information matrices for both the conditional and the Bayesian formulation of the super-resolution problem To simplify the presentation, we study the one-dimensional (1-D) version of the problem. Where applicable, we denote the unknown motion for the 1-D case by the scalar translation parameter $v_k$. We show in Appendix I that the extension to 2-D is straightforward.

We note that much of the analysis is simplified by examining the problem in the Fourier domain. Furthermore, many of the matrices are diagonalized by the discrete Fourier transform, allowing very efficient numerical implementation. In fact, for the 2-D case, the size of the matrices would necessarily preclude numerical calculation of the bounds were it not for the sparsity associated with the matrices in the Fourier domain. To differentiate between the Fourier and spatial domains, we use a tilde as in $\tilde{\mathbf{x}}$ to indicate vectors and matrices in the Fourier domain.

Before we begin, we must address the problem of complex notation associated with the Fourier domain. Because the unknown image $\mathbf{x}$ is real, the Fourier transform possesses a symmetry about the DC component in the frequency domain [15], [16]. As such, we would need to incorporate a symmetry constraint on the unknown image $\tilde{\mathbf{x}}$ in the frequency domain. Instead of employing a constrained CR bound, we rewrite the unknown frequency domain vector in the following fashion. We separate the real and imaginary components for the *positive* spatial frequencies and stack them in a vector as in

$$\tilde{\mathbf{z}} = \begin{pmatrix} \mathrm{Re}\{\mathcal{Z}(\omega_i)\} \\ \mathrm{Im}\{\mathcal{Z}(\omega_i)\} \end{pmatrix}, \quad \omega_i = \frac{i2\pi}{N_H} \tag{15}$$

where $\mathcal{Z}(\omega)$ is the DFT of the sampled signal $z_0, z_1, \ldots$ (the components of $\mathbf{z}$) and $i = 0, \ldots, N_H/2$ for the real portion and $i = 1, \ldots, (N_H/2) - 1$ for the imaginary portion. Here, the $\omega_i$ terms indicate the spatial frequencies comprising the signal $\mathbf{z}$. Using this separated form for the spectral signal, the dimensions of the spectral vectors $\tilde{\mathbf{x}}, \tilde{\mathbf{z}}, \tilde{\mathbf{y}}$ are equal to those of their spatial counterparts $\mathbf{x}, \mathbf{z}, \mathbf{y}$.

Under our proposed separated spectral representation, the convolution operators $\mathbf{F}$ and $\mathbf{H}$ are block diagonalized in the frequency domain. Assuming that the PSF is shift invariant, the blur matrix in the Fourier domain is given by

$$\tilde{\mathbf{H}} = \begin{pmatrix} \mathrm{diag}\{\mathrm{Re}[\mathcal{H}(\omega_i)]\} & \mathrm{diag}\{\mathrm{Im}[\mathcal{H}(\omega_i)]\} \\ \mathrm{diag}\{\mathrm{Im}[\mathcal{H}(\omega_i)]\} & \mathrm{diag}\{\mathrm{Re}[\mathcal{H}(\omega_i)]\} \end{pmatrix}. \tag{16}$$

The function $\mathcal{H}(\omega_i)$ is commonly referred to as the optical transfer function (OTF) of the imaging system. We make the additional assumption that the PSF is symmetric whereby $\mathrm{Im}[\mathcal{H}(\omega_i)] = 0, \forall i$. In keeping with the characterization of $\tilde{\mathbf{H}}$ as a low-pass filter (LPF), we assume that the filter has unit gain at D.C. ($|\mathcal{H}(0)| = 1$), and that, for each spatial frequency, $|\mathcal{H}(\omega_i)| \leq 1$.

Correspondingly, the shift matrix $\tilde{\mathbf{F}}_k$ is given by

$$\tilde{\mathbf{F}}_k = \begin{pmatrix} \mathrm{diag}\{\cos(v_k\omega_i)\} & -\mathrm{diag}\{\sin(v_k\omega_i)\} \\ \mathrm{diag}\{\sin(v_k\omega_i)\} & \mathrm{diag}\{\cos(v_k\omega_i)\} \end{pmatrix}. \tag{17}$$

Finally, the downsampling matrix $\mathbf{D}$ has the following structure $\tilde{\mathbf{D}} = (1/M)\begin{pmatrix} \tilde{\mathbf{D}}_R & 0 \\ 0 & \tilde{\mathbf{D}}_I \end{pmatrix}$ where

$$\{\tilde{\mathbf{D}}_R\}_{i,k} = \begin{cases} 1, & i = k - 2aN_L, \quad a = 0, 1, \ldots \lfloor M/2 \rfloor \\ 1, & i = 2aN_L - k, \quad a = 1, 2, \ldots \lfloor M/2 \rfloor \\ 0, & \text{else} \end{cases} \tag{18}$$

$$\{\tilde{\mathbf{D}}_I\}_{i,k} = \begin{cases} 1, & i = k - 2aN_L, \quad a = 0, 1, \ldots \lfloor M/2 \rfloor \\ -1, & i = 2aN_L - k, \quad a = 1, 2, \ldots \lfloor M/2 \rfloor \\ 0, & \text{else} \end{cases} \tag{19}$$

This structure reflects the spectral aliasing or *folding* due to the downsampling operation. Fig. 1 shows an example graphical representation of the matrix $\tilde{\mathbf{D}}$ for $M = 3$. We note that the upsampling operation ($\mathbf{D}^T$ in the time domain) is denoted $\tilde{\mathbf{D}}^{\ddagger}$ in the Fourier domain and is defined as $\tilde{\mathbf{D}}^{\ddagger} \equiv M\tilde{\mathbf{D}}^T$. Such a

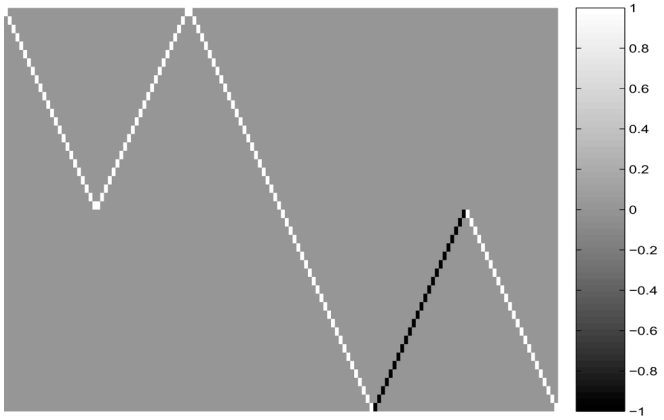


Fig. 1. Example image of the downsampling matrix $\bar{\mathbf{D}}$ for $M = 3$ showing the spectral folding of the frequency content due to downsampling.

definition is needed to ensure that $\tilde{\mathbf{D}}^{\ddagger}\tilde{\mathbf{D}}$ is an idempotent operator just as $\mathbf{D}^T\mathbf{D}$ is in the spatial domain. Given the structure associated with these matrices, we now proceed to derive the FIM.

### A. Fisher Information Matrix

We begin with the derivation of the FIM $\mathbf{J}$ due to the observed data used to construct the conditional CR bound. Using the invariance property, we only derive the FIM for the transformed model of (13). The conditional log-likelihood function for the observed data is given by

$$\ln p(\{\tilde{\mathbf{y}}_k\} \mid \tilde{\mathbf{z}}, \underline{\mathrm{v}}) = \frac{-1}{2\sigma^2}\sum_{k=0}^{K}(\tilde{\mathbf{y}}_k - \tilde{\mathbf{D}}\tilde{\mathbf{F}}(v_k)\tilde{\mathbf{z}})^T \times (\tilde{\mathbf{y}}_k - \tilde{\mathbf{D}}\tilde{\mathbf{F}}(v_k)\tilde{\mathbf{z}}) + \text{const.} \quad (20)$$

To derive the data FIM, we must compute the partial derivatives of the expected conditional log-likelihood function as follows:

$$-E_{\tilde{\mathbf{y}}\mid\tilde{\mathbf{z}},\underline{\mathrm{v}}}\left[\frac{\partial^2 \ln p(\{\tilde{\mathbf{y}}_k\}\mid\tilde{\mathbf{z}},\underline{\mathrm{v}})}{\partial\tilde{\mathbf{z}}^2}\right] = \frac{1}{\sigma^2}\left[\sum_{k=0}^{K}\tilde{\mathbf{F}}_k^T\tilde{\mathbf{D}}^{\ddagger}\tilde{\mathbf{D}}\tilde{\mathbf{F}}_k\right] \quad (21)$$

$$-E_{\tilde{\mathbf{y}}\mid\tilde{\mathbf{z}},\underline{\mathrm{v}}}\left[\frac{\partial^2 \ln p(\{\tilde{\mathbf{y}}_k\}\mid\tilde{\mathbf{z}},\underline{\mathrm{v}})}{\partial v_k^2}\right] = \frac{1}{\sigma^2}\left[\tilde{\mathbf{z}}^T\tilde{\mathbf{W}}^T\tilde{\mathbf{F}}_k^T\tilde{\mathbf{D}}^{\ddagger}\tilde{\mathbf{D}}\tilde{\mathbf{F}}_k\tilde{\mathbf{W}}\tilde{\mathbf{z}}\right] \quad (22)$$

$$-E_{\tilde{\mathbf{y}}\mid\tilde{\mathbf{z}},\underline{\mathrm{v}}}\left[\frac{\partial^2 \ln p(\{\tilde{\mathbf{y}}_k\}\mid\tilde{\mathbf{z}},\underline{\mathrm{v}})}{\partial v_k\partial\tilde{\mathbf{z}}}\right] = \frac{1}{\sigma^2}\left[\tilde{\mathbf{F}}_k^T\tilde{\mathbf{D}}^{\ddagger}\tilde{\mathbf{D}}\tilde{\mathbf{F}}_k\tilde{\mathbf{W}}\tilde{\mathbf{z}}\right]. \quad (23)$$

The matrix $\tilde{\mathbf{W}}$ arises from the fact that

$$\frac{\partial}{\partial v_k}\tilde{\mathbf{F}}(v_k) = \frac{\partial}{\partial v_k}\begin{pmatrix}\mathrm{diag}\{\cos(v_k\omega_i)\} & -\mathrm{diag}\{\sin(v_k\omega_i)\}\\ \mathrm{diag}\{\sin(v_k\omega_i)\} & \mathrm{diag}\{\cos(v_k\omega_i)\}\end{pmatrix} = \tilde{\mathbf{F}}(v_k)\begin{pmatrix}0 & -\mathrm{diag}\{\omega_i\}\\ \mathrm{diag}\{\omega_i\} & 0\end{pmatrix} = \tilde{\mathbf{F}}(v_k)\tilde{\mathbf{W}}. \quad (24)$$

The matrix $\tilde{\mathbf{W}}$, therefore, corresponds to a Fourier representation of a derivative operator in the spatial domain.

To simplify the representation, we define the following terms. First, we define $\tilde{\mathbf{d}} \equiv \tilde{\mathbf{W}}\tilde{\mathbf{z}}$, which is to say that $\tilde{\mathbf{d}}$ represents the sampled first derivative of the signal $\tilde{\mathbf{z}}$. Second, we define $\tilde{\mathbf{Q}}_k \equiv \tilde{\mathbf{F}}_k^T\tilde{\mathbf{D}}^{\ddagger}\tilde{\mathbf{D}}\tilde{\mathbf{F}}_k$. We note $\tilde{\mathbf{Q}}_k$ has the property of being a projection operator. The matrix $\tilde{\mathbf{Q}}_k$ can be expressed as a linear combination of matrices as

$$\tilde{\mathbf{Q}}_k = \frac{1}{M}\left(\mathbf{I} + \sum_{m=1}^{M-1}[\mathbf{\Lambda}_m\cos(m\phi_k) + \mathbf{\Gamma}_m\sin(m\phi_k)]\right) \quad (25)$$

where $\phi_k = (v_k 2\pi)/(M)$. The term $\phi_k$ can be thought of as the sampling phase offset between the first and the $k$th measured low-resolution image. Meanwhile, the matrices $\mathbf{\Lambda}_m$ and $\mathbf{\Gamma}_m$ are all symmetric matrices with zeros along the diagonal, representing the aliased portions of the folded spectrum due to the downsampling. As we see, the information content present in the signal is more directly dependent on the sampling phase offset $\phi_k$ as opposed to the translation $v_k$. In other words, the FIM is a periodic function of each of the translations $v_k$ with a period of $M$ pixels. This reflects the periodicity assumption about the original signal $\tilde{\mathbf{x}}$.

We observe that the FIM can be expressed as

$$\mathbf{J} = \begin{pmatrix}\tilde{\mathbf{A}} & \tilde{\mathbf{B}}\\ \tilde{\mathbf{B}}^T & \tilde{\mathbf{C}}\end{pmatrix} = \frac{1}{\sigma^2}\left(\begin{array}{c|ccc}\sum_{k=0}^{K}\tilde{\mathbf{Q}}_k & \tilde{\mathbf{Q}}_1\tilde{\mathbf{d}} & \cdots & \tilde{\mathbf{Q}}_K\tilde{\mathbf{d}}\\ \hline (\tilde{\mathbf{Q}}_1\tilde{\mathbf{d}})^T & \tilde{\mathbf{d}}^T\tilde{\mathbf{Q}}_1\tilde{\mathbf{d}} & & 0\\ \vdots & & \ddots & \\ (\tilde{\mathbf{Q}}_K\tilde{\mathbf{d}})^T & 0 & & \tilde{\mathbf{d}}^T\tilde{\mathbf{Q}}_K\tilde{\mathbf{d}}\end{array}\right). \quad (26)$$

The derivation of these terms for the 2-D case can be found in Appendix I. From this definition of the FIM, we obtain the conditional Schur information complements

$$\mathbf{S}_{\tilde{\mathbf{z}}} = \tilde{\mathbf{A}} - \tilde{\mathbf{B}}\tilde{\mathbf{C}}^{-1}\tilde{\mathbf{B}}^T \quad (27)$$

$$\mathbf{S}_{\underline{\mathrm{v}}} = \tilde{\mathbf{C}} - \tilde{\mathbf{B}}^T\tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}}. \quad (28)$$

These matrices are used to compute the conditional form of the scalar CR bound $T(\tilde{\mathbf{z}})$ in Section IV-A–D.

### B. Bayesian Information Matrices

In some instances, one may have certain prior information about the unknown parameters of interest. In our case, we assume that that the superresolution practitioner has prior information about the signal of interest $\tilde{\mathbf{x}}$, but has no information (flat prior) about the unknown motion parameters $\underline{\mathrm{v}}$. As we will show in the following section, such prior information is often critical during the image restoration phase of super-resolution.

Perhaps the most common form of prior information is a Gaussian distribution over the space of unknown images $\tilde{\mathbf{x}}$ [2], [14]. Thus, the unknown image comes from the distribution $\tilde{\mathbf{x}} \sim \mathcal{N}(\mu_{\tilde{\mathbf{x}}}, (1/\lambda)\mathbf{R}_{\tilde{\mathbf{x}}})$, where $\mu_{\tilde{\mathbf{x}}}$ is the mean image (typically zero) with a covariance $\mathbf{R}_{\tilde{\mathbf{x}}}$ and $\lambda$ is a parameter capturing the overall confidence in the prior knowledge. In an algorithmic setting, $\lambda$ is often used as a tuning parameter to control the strength of the prior information and, hence, its effect on the final estimate.

Typically, the image signal is assumed to have a diagonal covariance matrix $\mathbf{R}_{\tilde{\mathbf{x}}}$ (reflecting the stationarity of the random process) of the form

$$\mathbf{R}_{\tilde{\mathbf{x}}} = \begin{pmatrix} \mathrm{diag}\{[\mathcal{X}(\omega_i)]\} & 0 \\ 0 & \mathrm{diag}\{[\mathcal{X}(\omega_i)]\} \end{pmatrix} \tag{29}$$

where $\mathcal{X}(\omega)$ is the power spectral density (the Fourier transform of the image autocorrelation function) for the signal $\mathbf{x}$. Prior information of this sort stems from some physical property relating to the spatial smoothness of the signal. In the Fourier domain, a natural measure of smoothness is given by $\mathcal{X}(\omega) \approx (1/|\omega|^{\eta})$ where $\eta$ defines the global smoothness of the signal [16].[4] As one would expect, prior information about the unknown image $\tilde{\mathbf{x}}$ naturally offers information about $\tilde{\mathbf{z}}$. Specifically, we see that a Gaussian prior on $\tilde{\mathbf{x}}$ corresponds to a Gaussian prior on $\tilde{\mathbf{z}}$ as $\tilde{\mathbf{z}} \sim \mathcal{N}(\mu_{\tilde{\mathbf{z}}}, (1/\lambda)\mathbf{R}_{\tilde{\mathbf{z}}}) = \mathcal{N}(\tilde{\mathbf{H}}\mu_{\tilde{\mathbf{x}}}, (1/\lambda)\tilde{\mathbf{H}}\mathbf{R}_{\tilde{\mathbf{x}}}\tilde{\mathbf{H}}^T)$. For the remainder of the paper, we assume that the mean of the signal is zero $\mu_{\tilde{\mathbf{x}}} = 0$. This form of prior information is commonly utilized in the literature motivates the regularization penalty term in MAP estimators.

Typically, iterative super-resolution algorithms operating in the spatial domain use a finite-impulse response (FIR) filter to approximate $\mathbf{R}_{\tilde{\mathbf{x}}}^{-1}$. The most common filter used to regularize the image estimates is the Laplacian filter whose 1-D analog is given by $[-1, 2, -1]$ which is the first-order approximation of $\mathbf{R}_{\tilde{\mathbf{x}}}^{-1}$ when $\eta = 2$. In practice, higher order filters can be, but rarely are, used to more effectively incorporate prior information. Throughout what follows, we assume that the Laplacian filter approximation is used wherever prior information is assumed.

This type of prior information is generic in the sense that it can be applied to a large class of images. Unfortunately, the generality of such prior information ultimately reduces its effectiveness in improving performance. Ideally, the practitioner of super-resolution may be able to ascertain more precise information for a particular application. In such situations, statistical properties about a certain class of images can be *learned* from large data sets providing very useful information. For instance, the authors of [5], [18], and [19] show examples of incorporating learning-based priors into super-resolution for different classes of images.

Since the log of the Gaussian prior distribution $l(\tilde{\mathbf{z}})$ is

$$\ln p(\tilde{\mathbf{z}}) = -\frac{\lambda}{2}(\tilde{\mathbf{z}} - \mu_{\tilde{\mathbf{z}}})^T \mathbf{R}_{\tilde{\mathbf{z}}}^{-1}(\tilde{\mathbf{z}} - \mu_{\tilde{\mathbf{z}}}) - \ln\left(\left[\frac{2\pi}{\lambda}\right]^{\frac{N}{2}} |\mathbf{R}_{\tilde{\mathbf{z}}}|^{\frac{1}{2}}\right) \tag{30}$$

the prior information matrix term relating to the image is

$$-E_{\tilde{\mathbf{z}}}\left[\frac{\partial^2 l(\tilde{\mathbf{z}})}{\partial \tilde{\mathbf{z}}^2}\right] = \lambda \mathbf{R}_{\tilde{\mathbf{z}}}^{-1}. \tag{31}$$

In other words, the information depends only on the covariance term $\mathbf{R}_{\tilde{\mathbf{z}}}$. Finally, the prior information matrix is given by

$$\mathbf{J}_p = \begin{pmatrix} \lambda \mathbf{R}_{\tilde{\mathbf{z}}}^{-1} & 0 \\ 0 & 0 \end{pmatrix}. \tag{32}$$

[4]For natural images, the smoothness is modeled by $\eta = 2$. The foundations of this prior information can be traced to physical properties inherent to natural scenes [17].

Again, the off diagonal terms are zero since there is a noninformative prior on the motion parameters. We observe that the prior information matrix offers more information for the higher spatial frequencies helping to remedy the information lost due to the blurring degradation of the imaging system.

To compute the Bayesian data information matrix $\mathbf{J}_d$, we simply take the expected value of the FIM of (26). Assuming that the signal $\tilde{\mathbf{z}}$ is zero-mean, the elements of $\mathbf{J}_d$ are

$$E_{\tilde{\mathbf{z}}}[\tilde{\mathbf{A}}] = \tilde{\mathbf{A}} \tag{33}$$

$$E_{\tilde{\mathbf{z}}}[\tilde{\mathbf{B}}] = 0 \tag{34}$$

$$E_{\tilde{\mathbf{z}}}[\tilde{\mathbf{C}}] = \frac{1}{\sigma^2}\mathrm{diag}\{E_{\tilde{\mathbf{z}}}[\tilde{\mathbf{d}}^T \tilde{\mathbf{Q}}_k \tilde{\mathbf{d}}]\} \tag{35}$$

$$= \frac{1}{\sigma^2}\mathrm{diag}\left\{E_{\tilde{\mathbf{z}}}\left[\mathrm{tr}\left(\tilde{\mathbf{D}}\tilde{\mathbf{F}}_k \tilde{\mathbf{d}}\tilde{\mathbf{d}}^T \tilde{\mathbf{F}}_k^T \tilde{\mathbf{D}}^{\dagger}\right)\right]\right\} \tag{36}$$

$$= \frac{1}{\sigma^2}\mathrm{diag}\left\{\frac{1}{\lambda}\mathrm{tr}\left(\tilde{\mathbf{D}}\tilde{\mathbf{F}}_k \tilde{\mathbf{W}}\mathbf{R}_{\tilde{\mathbf{z}}}\tilde{\mathbf{W}}^T \tilde{\mathbf{F}}_k^T \tilde{\mathbf{D}}^{\dagger}\right)\right\} \tag{37}$$

$$= \frac{1}{\sigma^2\lambda}\mathrm{tr}(\tilde{\mathbf{W}}\mathbf{R}_{\tilde{\mathbf{z}}}\tilde{\mathbf{W}}^T)\mathbf{I}_K. \tag{38}$$

The last step assumes a diagonal covariance matrix $\mathbf{R}_{\tilde{\mathbf{z}}}$ of the form shown in (29). Finally, the Bayesian data information matrix is

$$\mathbf{J}_d = \begin{pmatrix} \tilde{\mathbf{A}} & 0 \\ 0 & \frac{1}{\sigma^2\lambda}\mathrm{tr}(\tilde{\mathbf{W}}\mathbf{R}_{\tilde{\mathbf{z}}}\tilde{\mathbf{W}}^T)\mathbf{I}_K \end{pmatrix}. \tag{39}$$

Thus, the Schur information complements for the Bayesian information matrix $\mathbf{J}_d + \mathbf{J}_p$ are

$$\mathbf{S}_{\tilde{\mathbf{z}}} = \tilde{\mathbf{A}} + \lambda \mathbf{R}_{\tilde{\mathbf{z}}}^{-1} \tag{40}$$

$$\mathbf{S}_{\underline{\mathrm{v}}} = \frac{1}{\sigma^2\lambda}\mathrm{tr}(\tilde{\mathbf{W}}\mathbf{R}_{\tilde{\mathbf{z}}}\tilde{\mathbf{W}}^T)\mathbf{I}_K. \tag{41}$$

These matrices are used to construct the scalar CR bound used in Section IV-E.

It is interesting to note that in the Bayesian case, increasing the variance of the signal distribution (decreasing $\lambda$), actually improves the bound on registration. In other words, an image signal with greater variation has more texture with which to register the images.

## IV. Analysis of the CR Bounds

In this section, we explore the various aspects of the CR bounds as they relate to super-resolution performance. Specifically, we study the relationship between image content, noise power, the imaging system's PSF, and the collection of motion vectors. To illuminate the complex nature of this relationship, we study the problem in the context of a series of questions regarding individual parameters. We then evaluate the CR bounds numerically restricting our attention to a single parameter of interest at a time to maintain clarity of presentation. In Section IV-A–C, we study the conditional CR bounds revealing the interplay between the motion vectors $\underline{\mathrm{v}}$ and image reconstruction. In Section IV-D, we study the subsequent problem of restoring the high-resolution image from the reconstructed image. In each of these sections, we encounter conditions in which the FIM becomes singular, and, hence, prior information must be applied to the estimation problem. In Section IV-E, we

study the performance bound improvements resulting from incorporating prior information into a Bayesian estimation framework. Finally, in Section IV-F, we compare the actual performance of a 2-D estimator with the 2-D conditional CR bound.

### A. What Are Good Sets of Motion Vectors?

The key component enabling the entire super-resolution process is the motion between the observed low-resolution images. In this section, we study the conditional performance limit on super-resolution as it relates to the set of motion vectors $\{v_k\}$ or, correspondingly, the set of sampling offsets $\{\phi_k\}$. To do so, we study a commonly used optimistic lower bound on image reconstruction performance which assumes the motion vectors are known exactly. Then, we find a set of sampling offsets which is optimal with respect to this weak bound.

Using the inverse of the Schur information complement (27), we observe that the MSE is bounded by

$$\begin{aligned}\mathbf{S}_{\tilde{\mathbf{z}}}^{-1} &= (\tilde{\mathbf{A}} - \tilde{\mathbf{B}}\tilde{\mathbf{C}}^{-1}\tilde{\mathbf{B}}^T)^{-1} \\ &= \tilde{\mathbf{A}}^{-1} + \tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}}\mathbf{S}_{\underline{\mathrm{v}}}^{-1}\tilde{\mathbf{B}}^T\tilde{\mathbf{A}}^{-1}. \end{aligned} \tag{42}$$

The first term on the right-hand side relates to the image reconstruction performance bound when the estimator has complete knowledge of the motion. The second term relates to the uncertainty introduced by having to estimate the unknown motion vectors. From this, we construct a weak lower bound on reconstruction performance by ignoring the second term giving

$$T_{\mathrm{weak}}(\tilde{\mathbf{z}}) = \left(\frac{1}{N_H}\mathrm{tr}(\tilde{\mathbf{A}}^{-1})\right)^{\frac{1}{2}}. \tag{43}$$

Even as a weak lower bound, much can be learned about the super-resolution problem through analysis of $T_{\mathrm{weak}}(\tilde{\mathbf{z}})$.

The matrix $\tilde{\mathbf{A}}$ has the general interpretation of representing the *amount* of observed data at each high-resolution pixel. We use the term *amount* rather than the number of observations because when the sampling offset falls in between two *grid points* (i.e., not an integer), the pixel measurement is spread across the local pixels. By grid points, we refer to the common terminology used to describe the $M$, (or $M \times M$ for 2-D), sample locations corresponding to integer shifts of the high-resolution image [14]. The condition number of $\tilde{\mathbf{A}}$ is related to the performance of signal reconstruction from interlaced sampling. This problem has been well studied in the signal processing community. For instance, [20] analyzes the stability of reconstruction for a given set of sampling offsets. In [20] it was shown that for certain collections of sampling offsets, the condition number for $\tilde{\mathbf{A}}$, and, hence, $\mathrm{tr}(\tilde{\mathbf{A}}^{-1})$ could grow quite large. Furthermore, the authors show that the ideal sampling offsets corresponds to translations that are uniformly distributed on the high-resolution *grid*.

We now introduce the notion of *evenly* spaced motion vector sets and show the general property that $T_{\mathrm{weak}}(\tilde{\mathbf{z}})$ is minimized when $\underline{\mathrm{v}}$ is an evenly spaced motion vector set. Furthermore, we see that the set of integer translations, observed to be optimal in [20], is but one example in the universe of evenly spaced motion vector sets. To define the set of evenly spaced sampling offsets, we first define the unit vector $\mathbf{u}_k = [\cos(\phi_k), \quad \sin(\phi_k)]^T$ constructed for each shift $v_k$ in $\underline{\mathrm{v}}$. A set $\underline{\mathrm{v}}$ is defined as being evenly spaced if the sum of these unit vectors is zero, or $\sum_{k=0}^{K}\mathbf{u}_k = 0$. For example, if $M = 2$, then both $\underline{\mathrm{v}} = [0, 0, 1, 1]$ and $\underline{\mathrm{v}} = [0, 0.5, 1, 1.5]^T$ are examples evenly spaced motion vector sets.

To see the optimality of evenly spaced motion vector sets with respect to $T_{\mathrm{weak}}(\tilde{\mathbf{z}})$, we rely on the inequality

$$\mathrm{tr}(\tilde{\mathbf{A}}^{-1}) \geq \sum_i \frac{1}{\{\tilde{\mathbf{A}}\}_{ii}} \tag{44}$$

which applies to all symmetric matrices [21]. From (26), we see that $\tilde{\mathbf{A}} = (1/\sigma^2)\sum_{k=0}^{K}\tilde{\mathbf{Q}}_k$. Since every term along the diagonal of the $N_H \times N_H$-dimensional matrix $\tilde{\mathbf{Q}}_k$ is $(1/M)$, we have

$$\begin{aligned}\sum_i \frac{1}{\{\tilde{\mathbf{A}}\}_{ii}} &= \sum_i \frac{\sigma^2}{\left(\sum_{k=0}^{K}\{\tilde{\mathbf{Q}}_k\}_{ii}\right)} = \sum_i \frac{M\sigma^2}{K+1} \\ &= \frac{\sigma^2 N_H M}{K+1} \end{aligned} \tag{45}$$

which is independent of $\underline{\mathrm{v}}$. We now show that $\mathrm{tr}(\tilde{\mathbf{A}}^{-1})$ satisfies the equality condition of (44) when $\underline{\mathrm{v}}$ is evenly spaced, thereby proving that $T_{\mathrm{weak}}(\tilde{\mathbf{z}})$ is optimized.

Using (25), we note that, when $\underline{\mathrm{v}}$ is evenly spaced, the matrix $\tilde{\mathbf{A}}$ is given by

$$\begin{aligned}\tilde{\mathbf{A}} &= \frac{1}{M}\sum_{k=0}^{K}\left(\mathbf{I} + \sum_{m=1}^{M-1}\left[\mathbf{\Lambda}_m\cos(m\phi_k) + \mathbf{\Gamma}_m\sin(m\phi_k)\right]\right) \\ &= \frac{K+1}{\sigma^2 M}\mathbf{I} \end{aligned} \tag{46}$$

where the trigonometric terms cancel because the sampling offsets are evenly spaced. Thus, the trace of the $N_H \times N_H$-dimensional matrix $\tilde{\mathbf{A}}^{-1}$ is equal to $(\sigma^2 N_H M)/(K+1)$ satisfying the equality of (44).

We have shown that a set of evenly spaced motion vectors $\underline{\mathrm{v}}$ are ideal from an image reconstruction perspective assuming the motions are known perfectly. This satisfies the intuition that the low-resolution images should adequately sample the high-resolution image. In what follows, we show that a subset of evenly spaced motion sets, we call *equally* spaced motion sets, defined by $v_k = (kM)/(K+1)$, offer nearly optimal performance bound characteristics with respect to the conditional CR bound.

### B. How Is Performance Affected by the Need to Estimate Both $\underline{\mathrm{v}}$ and $\tilde{\mathbf{z}}$

We now study the conditional performance bound for the real scenario when the motion vectors must be estimated from the low-resolution image data. To do so, we must analyze the degradation in performance due to the second term $\tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}}\mathbf{S}_{\underline{\mathrm{v}}}^{-1}\tilde{\mathbf{B}}^T\tilde{\mathbf{A}}^{-1}$ in (42). At this point, we resort to numerical evaluation of the CR bound.

In the previous section, we showed the optimality of equally spaced sampling offsets for image reconstruction when the motion vectors are known. It was shown that the weak performance bound for such a scenario was given by $T_{\mathrm{weak}}(\tilde{\mathbf{z}}) = \sigma\sqrt{(M)/(K+1)}$. One such clue as to the weakness of this bound relates to its independence of the image size $N_H$. As one would expect, when performing image registration, the size of the image $N_H$ plays a crucial role in registration performance. To give an idea of the overall degradation in performance due to uncertainty about the motion parameters, we compare the numerical evaluation of the conditional CR bound $T(\tilde{\mathbf{z}})$ (when

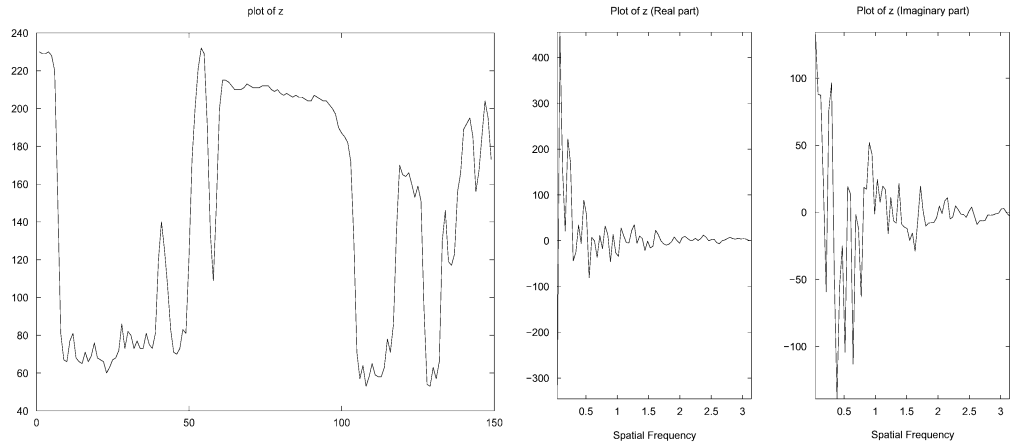


Fig. 2. Shown is a plot of the 150-pixel experimental test signal used throughout our conditional 1-D CRB simulations. (Left) The signal is shown in the spatial domain and (right) in the separated frequency domain.

motion vectors must be estimated) with the optimistic bound $T_{\text{weak}}(\tilde{\mathbf{z}})$. This comparison reveals the loss in performance that can be expected when the motion vectors must be estimated from the observed data. We evaluate the complete bound $T(\tilde{\mathbf{z}})$, using cropped versions of the test signal shown in Fig. 2 to explore the effect of signal size on reconstruction performance. The signal was truncated to the first 60, 90, and 120 pixels of the original 150-pixel signal. Fig. 3 compares these two performance bounds for equally spaced motions $v_k = (kM)/(K+1)$ assuming $M = 4, \sigma^2 = 1$.

We see that the performance loss due to registration uncertainty is quite significant, especially for smaller signals. As signals increase in size, the performance improves up to a saturation point at which we observe the unavoidable loss in performance resulting from the need to estimate the motion vectors. Relatively speaking, a conservative expected performance degradation ranges from 10% to 25%, as the number of available frames $K+1$ increase. In other words, when a large number of frames are observed, using the simple bound which assumes motions are known can be overly optimistic by as much as 25%.

We now show that, in general, the CR bound has a signal-dependent, spatially varying character which depends on two terms in (42). First, when the set of sampling offsets deviate from the equally spaced condition, the per-pixel variance bound takes on a periodic structure indicative of the *amount* of measurement at the high-resolution grid points. This component is independent of the signal $\mathbf{z}$ and depends only on the set of sampling offsets $\{\phi_k\}$ through the first term $\tilde{\mathbf{A}}^{-1}$. Second, the term $\tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}}\mathbf{S}_{\underline{\mathrm{v}}}^{-1}\tilde{\mathbf{B}}^T\tilde{\mathbf{A}}^{-1}$ captures the uncertainty associated with the estimation of the unknown sampling offsets. This term shows that reconstruction performance is degraded in regions with large spatial derivatives. A simple example of such behavior is derived for $M = 1$ in Appendix III, where it is shown that, regardless of $\underline{\mathrm{v}}$, the second term is given by

$$\tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}}\mathbf{S}_{\underline{\mathrm{v}}}^{-1}\tilde{\mathbf{B}}^T\tilde{\mathbf{A}}^{-1} = \frac{K}{K+1}\frac{\tilde{\mathbf{d}}\tilde{\mathbf{d}}^T}{\tilde{\mathbf{d}}^T\tilde{\mathbf{d}}}. \tag{47}$$

In other words, the variance bound in estimating a particular pixel depends on that pixel's relative gradient magnitude or texture.

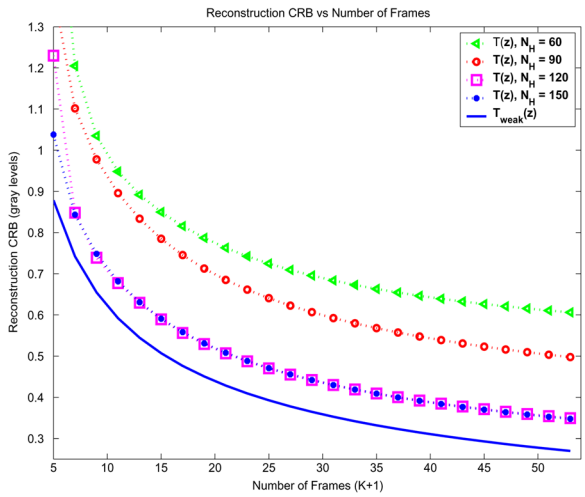


Fig. 3. Shown are the numerically computed conditional reconstruction CR Bound $T(\bar{\mathbf{z}})$ for cropped versions of (symbols) the test signal and (solid line) the optimistic lower bound $T_{\text{weak}}(\bar{\mathbf{z}})$ plotted as a function of the number of frames $K+1$. The cropped signals are truncated to the first 60, 90, 120, and 150 pixels of the test signal. The bounds are computed for equally spaced motions $v_k = (kM)/(K+1), M = 4$, and $\sigma^2 = 1$. (Color version available online at http://ieeexplore.ieee.org.)

As an example, the graph of Fig. 4 shows the spatial domain per-pixel variance bound for estimating the gray levels of the test signal. The bound was calculated assuming $K+1 = 4$ measured low-resolution images with the set of translations $\underline{\mathrm{v}} = [0, 0.5, 1, 2]^T$ (not evenly spaced), a downsampling factor of $M = 3$, and a noise power of $\sigma^2 = 1$. Here, we show the bound in the spatial domain to simplify its interpretability. The per-pixel variance bound has the two distinct characteristics. First, we see the sawtooth-like periodic function stemming from $\tilde{\mathbf{A}}^{-1}$. Second, we observe the spikes in variance bound resulting from the $\tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}}\mathbf{S}_{\underline{\mathrm{v}}}^{-1}\tilde{\mathbf{B}}^T\tilde{\mathbf{A}}^{-1}$ term. Note that the spikes in the bound correspond to the locations of the "edges" or high-frequency detail in the original spatial signal $\mathbf{z}$. Essentially, this reflects the intuitive observation that errors in motion estimation will be most detrimental to image reconstruction in highly textured or high spatial frequency areas. For example, poor registration during super-resolution restoration causes an

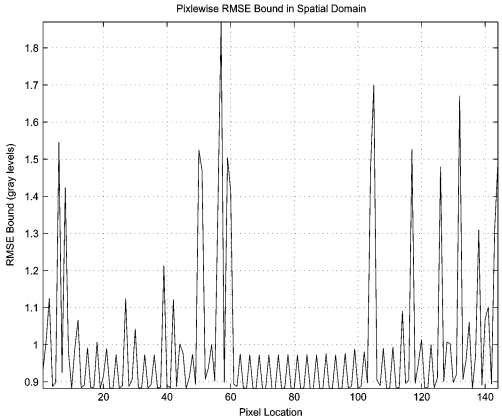


Fig. 4. Shown are the square root of the terms along the diagonal of $\mathbf{S}_{\mathbf{z}}^{-1}$ representing the conditional RMSE bound for each pixel location of the test signal. The bound was computed assuming a downsampling factor $M = 3$, $K+1 = 4$ low-resolution frames with shifts of $\underline{\mathrm{v}} = [0, 0.5, 1, 2]^T$, and $\sigma^2 = 1$. We observe that the pixel performance varies according to the sampling offsets as well as the edges in the signal.

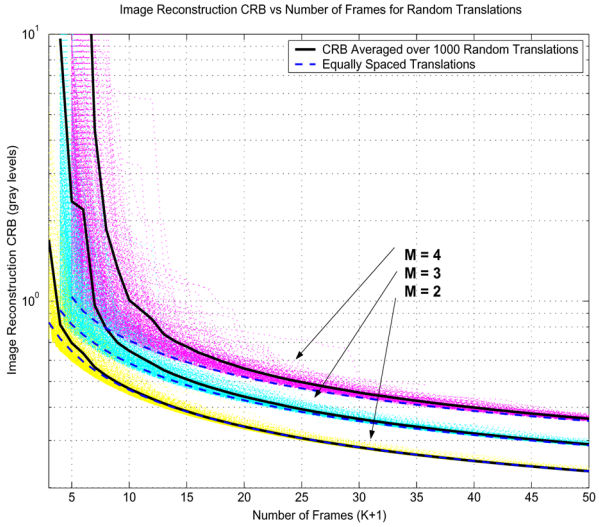


Fig. 5. Each scatter point represents the numerically computed conditional CR bound $T(\bar{\mathbf{z}})$ for a randomly selected set of translations $\underline{\mathrm{v}}$ from a uniform distribution over the range $[0, M]^K$. The reconstruction bound was computed for downsampling factors of $M = 2, 3, 4$, and a noise power of $\sigma^2 = 1$. The solid lines represent the average value of $T(\bar{\mathbf{z}})$ over the 1000 random sets of motion vectors drawn for each value of $K + 1$ and the dashed lines show the value of $T(\bar{\mathbf{z}})$ for equally spaced translations. (Color version available online at http://ieeexplore.ieee.org.)

edge-like feature to be distorted, exhibiting sawtooth type artifacts [1], [22]. This presents an interesting tradeoff in that the very image content which is easiest to register (highly textured) is also the content which is most prone to errors in the reconstruction/restoration.

In some cases, the need to produce unbiased estimates of the motion vectors $\underline{\mathrm{v}}$ *and* the image $\tilde{\mathbf{z}}$ in joint fashion makes the problem drastically harder as seen by a degradation of the performance bound. To see an example of this behavior, we need look no further than the canonical example presented in the super-resolution literature. Perhaps the most common experimental setup in the super-resolution literature is that of a blurry image $\mathbf{z}$, downsampling by a factor $M$ with integer motions $\underline{\mathrm{v}} = [0, 1, \ldots, M-1]$ to produce $K+1 = M$ low-resolution frames [14]. While this was shown earlier to be favorable from a signal reconstruction perspective (assuming the translations were known) [20], when the translations must also be estimated from the data, the FIM is, in fact, singular. We show this in Appendix II. In such situations, an unbiased estimator has insufficient information to both estimate the motion parameters and reconstruct the image. When the FIM is singular, any unbiased estimator of the set of translations will necessarily have infinite variance, or any estimator with finite variance must contain bias [23]. This example clearly shows another weakness in using only $T_{\mathrm{weak}}(\tilde{\mathbf{z}})$ (known-motion case) to predict performance in super-resolution.

### C. How Do the Downsampling Factor $M$, and the Number of Frames $K+1$, Affect Performance?

As we have just shown, the conditional CR bound $T(\tilde{\mathbf{z}})$ for the joint estimation problem can be dramatically worse than the optimistic bound $T_{\mathrm{weak}}(\tilde{\mathbf{z}})$. Because of the complexity of the CR bound's dependence on the set of motion vectors $\underline{\mathrm{v}}$, we now resort to Monte Carlo (MC) simulations to understand the CR bound as it relates to the downsampling factor $M$ and the number of observed frames $K+1$.

We compute the conditional CR bound $T(\tilde{\mathbf{z}})$ for the signal in Fig. 2 with randomly drawn translation sets $\underline{\mathrm{v}}$ from a uniform distribution over $[0, M]^K$. For each value of $K$, $T(\tilde{\mathbf{z}})$ was evaluated numerically for each of 1000 random sets of motion vectors $\underline{\mathrm{v}}$. Again, the noise power is $\sigma^2 = 1$. Fig. 5 shows the computed performance bound $T(\tilde{\mathbf{z}})$ for these randomly drawn translations as the cloud of points. The solid line indicates the average of $T(\tilde{\mathbf{z}})$ over the random set for each value of $K+1$. The values of $T(\tilde{\mathbf{z}})$ are plotted on a logarithmic scale due to their high degree of variability. As a point of reference, the dashed lines indicate the bound $T(\tilde{\mathbf{z}})$ for the equally spaced translations. Fig. 5 offers some insight into the functional relationship between $T(\tilde{\mathbf{z}})$ and $\underline{\mathrm{v}}$. For instance, it shows that as the number of frames increases, the variability of $T(\tilde{\mathbf{z}})$ diminishes quite substantially. To summarize, if given a large enough collection of images with (uniformly) random offsets, the performance bound can be expected to be very close to the bound for equally spaced sampling offsets.

Another way to interpret the data shown in Fig. 5 is to plot the number of frames necessary to achieve a specific overall RMSE bound. To visualize the CR bound in this fashion, we choose a target value for the overall RMSE bound which we denote $T_{\mathrm{targ}}(\tilde{\mathbf{z}})$. Then, we ask the following question: "As a function of the number of frames $K+1$, what percentage of the values of $T(\tilde{\mathbf{z}})$ for the randomly chosen translations are at least as good as $T_{\mathrm{targ}}(\tilde{\mathbf{z}})$?" Such a value relates to the probability that the CR bound, and, hence, the predicted performance, will fall below a target RMSE for uniformly random motions. Because the dimensions of the signal depend on $M$, we would compare performance bound curves for a normalized value of $K$. As such, we plot our curves as a function of $(K+1)/(M)$. Fig. 6 shows the numerically evaluated probability that $T(\tilde{\mathbf{z}})$ is greater than the target values $T_{\mathrm{targ}}(\tilde{\mathbf{z}}) = 1$ and $T_{\mathrm{targ}}(\tilde{\mathbf{z}}) = 1/2$. Here, we see that there is an expected performance degradation (as indicated by the required amount of image data) associated with the

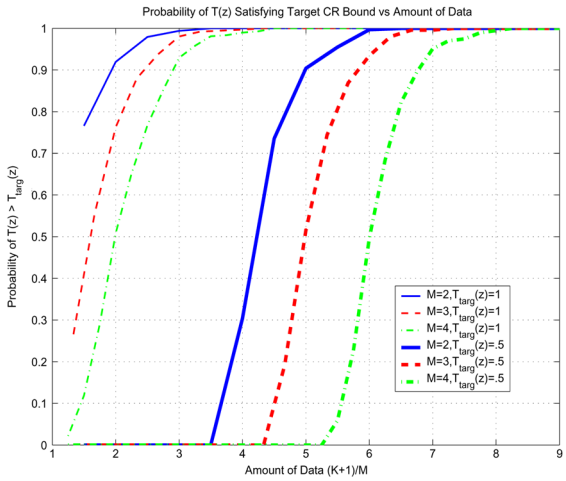


Fig. 6. Curves indicate the probability that the CR bound $T(\bar{\mathbf{z}})$ is less than the target RMSE performance of (thick lines) $T_{\text{targ}}(\bar{\mathbf{z}}) = 1$ and (thin lines) $T_{\text{targ}}(\bar{\mathbf{z}}) = 1/2$ as a function of the number of frames normalized by the downsampling factor $(K+1)/(M)$. The probabilities are computed using 1000 randomly drawn motion vectors for each point in the probability curves. The noise power is $\sigma^2 = 1$. (Color version available online at http://ieeexplore.ieee.org.)

downsampling factor $M$. Furthermore, this degradation is worsened as higher quality estimates are required.

Plots such as Fig. 6 offer the practitioner of super-resolution informative guidelines for choosing a particular resolution enhancement factor $M$ and number of frames $K+1$ to satisfy a particular target reconstruction MSE. For example, in [24], the idea was presented wherein a collection of images measuring a particular signal might be downsampled prior to image compression encoding with the understanding that super-resolution would be performed in a decoding step. The downsampling operation minimizes the bandwidth needed to transmit multiple measurements of a signal. For larger downsampling factors, however, greater number of frames are needed to achieve a target super-resolution performance. Plots such as Fig. 6 could help guide the choice of this downsampling ratio for a given set of constraints on channel bandwidth and computational resources.

### D. How Does the Imaging System's PSF Affect Performance?

In this section, we characterize the conditional performance bound associated with restoring the signal $\tilde{\mathbf{x}}$ from an estimate of the blurry signal $\hat{\tilde{\mathbf{z}}}$. This stage of the estimation problem relates to the well-known challenge of image deconvolution/deblurring. We will show that image restoration can represent the most difficult aspect of super-resolution from the perspective of performance bounds.

As we mentioned earlier, the CR bounds for image restoration depend on the reconstruction bounds as $\mathbf{S}_{\hat{\tilde{\mathbf{x}}}} = \tilde{\mathbf{H}}\mathbf{S}_{\tilde{\mathbf{z}}}\tilde{\mathbf{H}}$. Immediately, we see the instability of the restoration problem due to the poor conditioning of the $\tilde{\mathbf{H}}$ operator. The matrix $\tilde{\mathbf{H}}$ may be singular if there exists zero crossings in the OTF. In practice, the OTF of a given imaging system reflects a combination of several degradation processes [25] and it may be reasonable to expect zero crossings within the bandwidth of the signal $\tilde{\mathbf{x}}$. Furthermore, $\tilde{\mathbf{H}}^{-1}$ represents a high-pass type operator. As we have already shown, the CR bound on image reconstruction reveals that uncertainty is greatest at high-frequency locations (edges). Thus, the high-pass characteristic of $\tilde{\mathbf{H}}^{-1}$ is even more burdensome as it further amplifies this uncertainty.

For example, in the right graph of Fig. 7, we show the performance bound for different PSFs assuming that the signal of interest $\tilde{\mathbf{x}}$ was that of Fig. 2. The PSFs examined are represented by nine tap Gaussian filters with standard deviation of $w$ in pixels. The corresponding OTFs are shown in the left graph of Fig. 7 as a function of spatial frequency in units of radians per sample. The downsampling factor was $M = 2$ and $\sigma^2 = 1$. Again, we show the performance bound for equally spaced motions as the number of low-resolution measurements $K+1$ increases. The performance bounds are shown in log-scale because the degradation for different PSFs is so severe. This CR bound sensitivity underscores the ill-posed nature of image deconvolution. In fact, for the PSF with width $w = 1.0$, about 50 equally spaced low-resolution frames are needed to bring the per-pixel variance bound to under three gray levels (or three times the noise standard deviation $\sigma$). Such restoration performance is very poor indeed for real imaging applications.

When the OTF matrix $\tilde{\mathbf{H}}$ is singular, and, hence, the FIM $\tilde{\mathbf{H}}\tilde{\mathbf{S}}_{\mathbf{z}}\tilde{\mathbf{H}}$ is singular, it is essentially impossible to construct an unbiased estimate of the image $\tilde{\mathbf{x}}$. The singularities of the OTF effectively eliminate all information about specific frequencey spectral components of the signal. It is important to note at this point that in both in [5] and [6], the authors make an implicit connection between the PSF of the imaging system and the downsampling factor $M$. The implicit conclusion of [6] is that the factor $M = 1.6$ is the "fundamental limit under practical situations," where the PSF is assumed to be a boxcar function (loosely based on a pixel geometry with 100% fill factor) whose dimensions are dependent on the downsampling factor $M$. Such a conclusion is reasonable *only* in that the OTF corresponding to a boxcar PSF necessarily has zero crossings. Whether or not these zero crossings fall within the bandwidth of the unknown image ultimately dictates the performance associated with super-resolution. For example, in color imaging systems, the Bayer pattern inherent to single sensor imaging systems often contain sufficient aliasing to permit super-resolution well beyond a factor of 1.6 [1]. Furthermore, as we shall show in the next section, prior information can regularize the problem even when the PSF contains zeros within the signal bandwidth such that reasonable performance is achievable.

Because of the difficulty of image restoration, it is a practical necessity to tackle the problem of super-resolution with some prior knowledge about the unknown image. As also pointed out in [5], the restoration aspect of super-resolution may indeed be the critical information bottleneck to performance, requiring prior information to achieve reasonable performance. In the next section, we explore the role of prior information in the problem of super-resolution.

### E. How Does Prior Information Affect the Performance Bound?

Up to this point, we have studied the conditional CR bound on super-resolution under the assumption that no prior information was available. We have shown that were the translations properly spaced, and the PSF operator reasonably well conditioned,

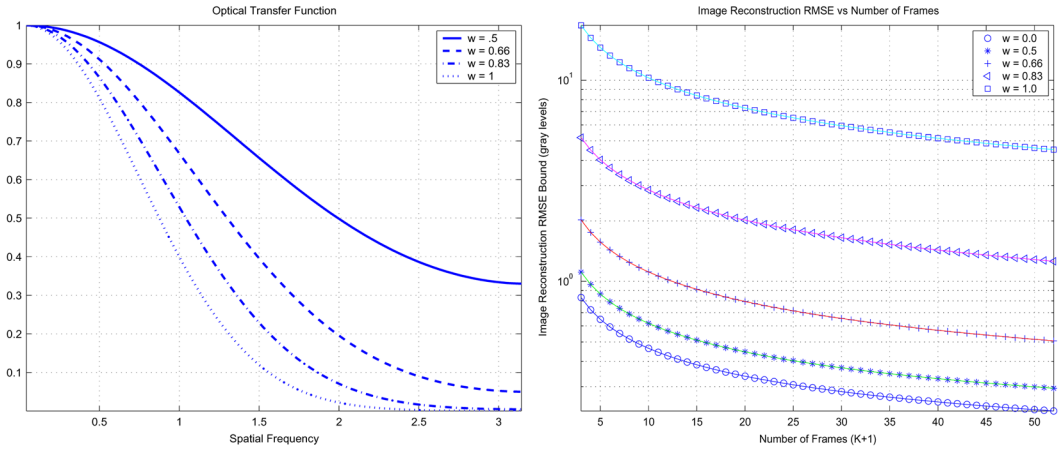


Fig. 7. Left graph shows the magnitude of the OTF $|\mathcal{H}(\theta)|$ for different widths $w$ of a Gaussian kernel PSF. The right graph shows the corresponding conditional image restoration CR bound $T(\mathbf{x})$ as a function of the number of frames $K+1$ for a downsampling factor of $M=2$, equally spaced motions $v_k=(kM)/(K+1)$, and a noise power of $\sigma^2=1$. (Color version available online at http://ieeexplore.ieee.org.)

super-resolution is possible without prior information. Yet, we have encountered several scenarios where the FIM exhibits singular behavior relating to either unfavorable motion vectors or zero-crossings in the OTF. Under these scenarios, the FIM becomes singular and super-resolution is possible only with prior information.

It was suggested in [5] that the performance limits to super-resolution can be "broken." The method used to overcome the performance limits was an effective form of prior information *learned* from training data. Our analysis agrees with the claims of [5] that prior information can "break" the performance limits in the sense that it improves the performance bound. Essentially, when the observation model is such that the conditional Schur information complement $\mathbf{S}_{\tilde{\mathbf{x}}}$ is singular, unbiased estimation with finite variance is impossible. By incorporating prior information, we can tradeoff biased estimation of a particular signal to gain finite variance. The goal in this Bayesian regime becomes that of finding estimators which perform well over the distribution of signals. As we now show, the Bayesian CR bound remains tractable even when the super-resolution using only the likelihood function would otherwise be extremely ill posed.

In our analysis thus far, we have encountered two scenarios in which the FIM becomes singular. The first is when the motion vector set $\underline{\mathrm{v}}$ insufficiently samples the high-resolution image causing $\tilde{\mathbf{A}}$ to be singular. In this case, prior information which helps condition the problem by improving the condition of $\mathbf{S}_{\tilde{\mathbf{z}}}$ as defined in (40). To see this effect, we conduct another set of MC simulations similar to those shown in Fig. 5. This time, however, we compute the Bayesian CR bound which incorporates different amounts of prior information parameterized by $\lambda$. Rather than plotting the point clouds, we only plot the average value of $T(\tilde{\mathbf{z}})$ for each value of $K+1$. Fig. 8 shows the average values of $T(\tilde{\mathbf{z}})$ when $M=4$ and $\sigma^2=1$. From Fig. 8, we see that the prior information is really only necessary when only a few frames are available. When the number of frames is low, the matrix $\tilde{\mathbf{A}}$ is much more likely to be singular. In these cases, the prior information is necessary to keep the variance under control. As the number of frames increase beyond a certain point (around ten frames), however, the prior information offers little

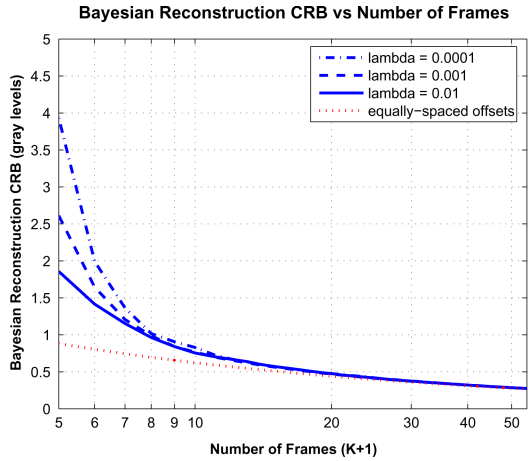


Fig. 8. Each point in the curves show the value of Bayesian CR bound $T(\tilde{\mathbf{z}})$ averaged over 1000 uniformly drawn motion vectors $\underline{\mathrm{v}} \in [0,M]^K$ for each value of $K+1$. The different line types indicate the amount of prior information used when computing $T(\tilde{\mathbf{z}})$. The light dotted line indicates the Bayesian CR bound $T(\tilde{\mathbf{z}})$ computed for equally spaced motions with $\lambda=10^{-8}$ (essentially no prior information). The bounds were computed for a downsampling factor of $M=4$ and a noise power of $\sigma^2=1$. (Color version available online at http://ieeexplore.ieee.org.)

advantage, as the matrix $\tilde{\mathbf{A}}$ becomes reasonably conditioned on average.

In the previous example, we observed that by increasing the number of frames, we could reasonably expect stable performance with little prior information. This is not the case when it comes to image restoration. As we showed in Section IV-D, the problem of image restoration may routinely be ill conditioned, if not ill posed. In these situations, increasing the number of frames will not improve the conditioning of the FIM. In such situations, prior information is required to perform image restoration with finite variance. Using (14), we see that

$$\mathbf{S}_{\tilde{\mathbf{x}}} = \tilde{\mathbf{H}}\left[\tilde{\mathbf{A}} + \lambda \mathbf{R}_{\tilde{\mathbf{z}}}^{-1}\right]\tilde{\mathbf{H}}^T \tag{48}$$

$$= \left[\tilde{\mathbf{H}}\tilde{\mathbf{A}}\tilde{\mathbf{H}}^T + \lambda \tilde{\mathbf{H}}\mathbf{R}_{\tilde{\mathbf{z}}}^{-1}\tilde{\mathbf{H}}^T\right] \tag{49}$$

$$= \left[\tilde{\mathbf{H}}\tilde{\mathbf{A}}\tilde{\mathbf{H}}^T + \lambda \mathbf{R}_{\tilde{\mathbf{x}}}^{-1}\right] \tag{50}$$

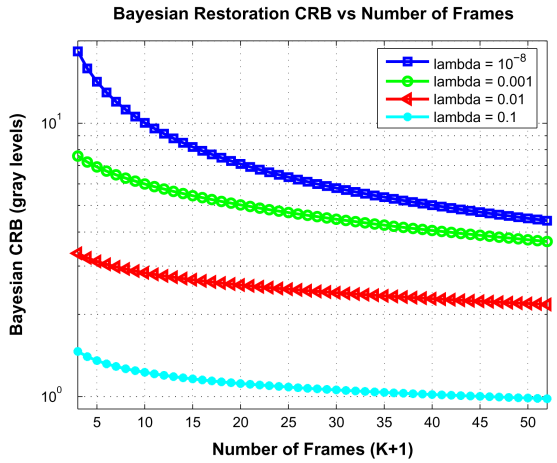


Fig. 9. Curves show the Bayesian restoration CR bound $T(\tilde{\mathbf{x}})$ for equally spaced motions $v_k = (kM)/(K+1)$ as a function of the number of frames $K+1$ for a downsampling factor of $M = 2$ and a noise power of $\sigma^2 = 1$. The PSF is assumed to be a Gaussian function with width $w = 1$, which is chosen such that the OTF does not reach zero but grows very small for high frequencies. The amount of prior information, parameterized by $\lambda$, greatly affects the predicted overall restoration performance. (Color version available online at http://ieeexplore.ieee.org.)

where the last step uses the relationship $\mathbf{R}_{\tilde{\mathbf{z}}} = \tilde{\mathbf{H}}\mathbf{R}_{\tilde{\mathbf{x}}}\tilde{\mathbf{H}}$. From (50), we see that even if the OTF has zero crossings (and, hence, $\tilde{\mathbf{H}}$ is singular) the Bayesian CR bound is well behaved as long as the additional prior information term $\mathbf{R}_{\tilde{\mathbf{x}}}$ is full rank.

As an illustration of the above observation, we repeat an experiment similar to that of Section IV-D for the PSF with width $w = 1.0$ while assuming a prior on the unknown image $\tilde{\mathbf{x}}$ of the form (29). Fig. 9 shows the improvement in the Bayesian CR bound resulting from the additional prior information. Without the prior information, the predicted performance was unacceptably poor, even with 50 low-resolution measurements. With even mild amounts of prior information, the performance bound falls into the range of reasonable performance ($<2\sigma$). For this reason, prior information is a necessity for the restoration stage of super-resolution.

### *F. What About 2-D Images?*

In this section, we compare the performance of an actual multiframe image reconstruction algorithm with the numerically evaluated deterministic CR bound for a 2-D test image. These simulations show that the conditional 2-D CR bound, as derived in Appendix I, offers reasonable prediction of estimator performance.

While there are a variety of multiframe image reconstruction algorithms in the literature [2], [4], [14], each of these algorithms round the motion vector estimates onto the nearest integer (a.k.a. grid location). The motion vectors in our simulations, however, do not fall onto these sampling grid points, necessitating a slightly different approach. The algorithm used in our experiments follows the standard, yet suboptimal, two-stage noniterative approach to multiframe image reconstruction similar to that proposed in [14]. First, the set of motion vectors are estimated in a pair wise fashion. Second, using the estimated motion vectors $\{\mathbf{v}_k\}$, the low-resolution frames are combined to form an estimate of the original image.

To register the low-resolution images, we implemented the relative phase algorithm of [26] which was specifically proposed to address the problem of registering a pair of aliased images. In deriving the algorithm, the authors make several heuristic observations which they use to motivate the application of a nonlinear weighting of zeros and ones (a mask) to the spectrum of the images prior to a phase-based registration algorithm. This masking prunes away the portions of the image spectrum where the negative effects of aliasing are assumed to worsen estimation performance. For our experiments, we used parameter settings recommended in [26] which have been shown to offer good performance for general image registration [27].

To reconstruct the original high-resolution image, we compute a direct estimate of the image as

$$\hat{\tilde{\mathbf{z}}} = \left(\sum_k \tilde{\mathbf{Q}}(\hat{v_k})\right)^{-1} \left(\sum_k \tilde{\mathbf{Q}}(\hat{v_k})^T \tilde{\mathbf{y}}_k\right). \tag{51}$$

Such an approach is a frequency-based generalization of the straightforward approach of [14] for noninteger motion vectors. The algorithm differs from [14] in that it is by no means a computationally efficient algorithm as it requires the explicit construction of extremely large, albeit sparse, matrices.

We perform our experiments using the $150 \times 150$ pixel Tree image shown in the left graph of Fig. 10. For our experiments, we construct low-resolution images using equally spaced translation on a 2-D grid. In other words, the translations vectors are defined by the set of ordered pairs $\{\ldots(kM/(K+1))\ldots\} \times \{\ldots(kM)/(K+1)\ldots\}$ We evaluated the reconstruction performance using $K+1 = 9, 16, 25, 36$ frames for a downsampling factor of $M = 2$. The right graph of Fig. 10 shows a scatterplot of the motion vectors. The reference frame used in the image registration process was the low-resolution image produced by the motion vector $\mathbf{v} = [0, 0]^T$.

We evaluated the algorithm's reconstruction performance for SNR values of 30, 40, and 50 dB. To evaluate the conditional RMSE performance $\overline{\text{rmse}}(\tilde{\mathbf{z}})$, the performance of the estimator was averaged over 500 MC simulation runs for each data point. The left graph of Fig. 11 compares the actual algorithm performance $\overline{\text{rmse}}(\mathbf{z})$ with the predicted CR bound $T(\mathbf{z})$ versus the number of equally spaced frames. The right graph of Fig. 11 shows the per-pixel RMSE performance at an SNR of 40 dB for $K+1 = 25$ frames. As predicted in Section IV-B, the estimator error is most severe near the edge locations in the Tree image.

From the left graph of Fig. 11, we observe that the CR bound predicts the actual estimator performance for SNR of 30 and 40 dB reasonably well. At 50 dB, however, the estimator performance departs from what the CR bound predicts. This is explained by the systematic bias associated with the registration algorithm [27]. To show this bias, Fig. 12 compares the registration performance $\overline{\text{rmse}}(\underline{\text{v}})$ with the CR bound $T(\underline{\text{v}})$ for the same experiment. There are two notable features of Fig. 12. First, the overall registration performance changes very little as a function of the number of frames. This is intuitive given that pairwise registration is performed independent of the reconstruction. Second, the actual performance flattens out even though the CR bound suggests improved performance for increasing SNR.

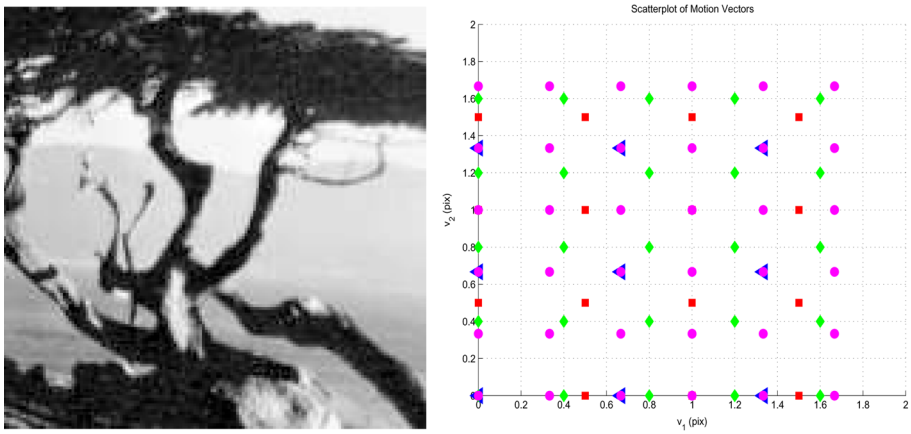


Fig. 10. Left image shows the $150 \times 150$ pixel Tree image used for the experiment. The graph on the right shows a scatter plot of the motion vectors for $K+1 = 9$ ($\triangleleft$), $K+1 = 16$ ($\square$), $K+1 = 25$ ($\diamond$), and $K+1 = 36$ ($\circ$). (Color version available online at http://ieeexplore.ieee.org.)

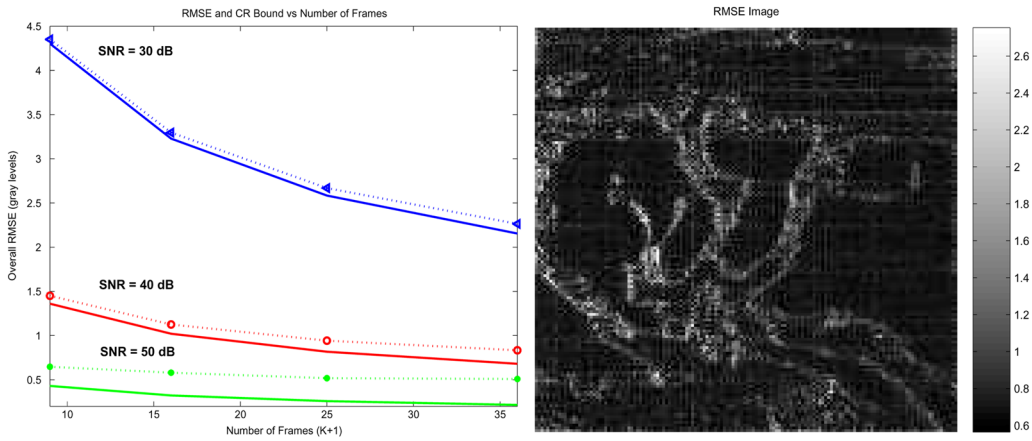


Fig. 11. Left graph compares the actual estimator reconstruction performance (symbols) $\overline{\mathrm{rmse}}(\bar{\mathbf{z}})$ and the corresponding conditional CR bounds (solid lines) $T(\bar{\mathbf{z}})$ as a function of the number of frames $K+1$. Each point represents the average over 500 MC simulations. The right graph depicts the actual estimator RMSE image over 500 MC runs for $K+1 = 25$ at 40 dB. (Color version available online at http://ieeexplore.ieee.org.)

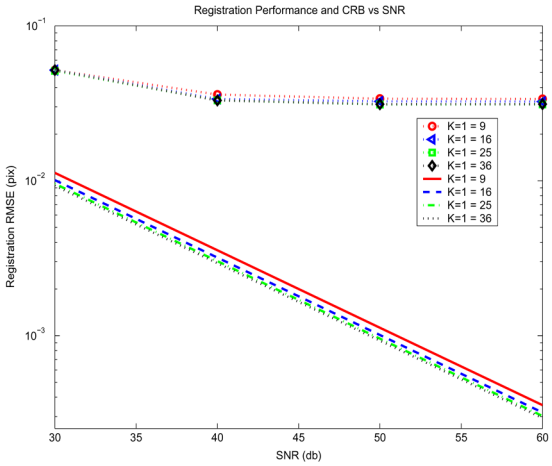


Fig. 12. Graph shows the (solid lines) CR bounds $T(\bar{\mathrm{v}})$ and (symbols) actual registration performance $\overline{\mathrm{rmse}}(\underline{\mathrm{v}})$ versus the SNR for the 2-D equally spaced motions with $K+1 = 9, 16, 25, 36$. (Color version available online at http://ieeexplore.ieee.org.)

This flattening out of the performance is indicative of the deterministic bias associated with the registration algorithm [27].

This registration bias prevents the reconstruction algorithm from achieving the CR bound for high SNR. At low SNR, the reconstruction error due to registration bias is masked by the error due to the additive measurement noise. As the SNR improves, however, the registration bias limits the final reconstruction performance causing edge artifacts as observed in the right graph of Fig. 11.

## V. Conclusion

In this paper, we have derived and explored the use of the CR inequality in bounding performance for the problem of super-resolution from a statistical perspective. Our analysis shows that there is no single bound which applies to all super-resolution operating scenarios as others have done in the past. Instead, we have shown that super-resolution performance depends on a complex relationship between measurement SNR, the number of observed frames, set of relative motions between frames, image content, and the imaging system's PSF. In addition, we presented an alternate representation of the problem in the Fourier domain which facilitates the numerical evaluation of the CR bound.

Our analysis offers practical new insight into the challenges of super-resolution. For instance, we showed that the degradation in super-resolution performance can be substantial when image motion must be estimated from the data, as opposed to

being known *a priori*. We showed that this degradation occurs most severely along edges within images. In addition, we showed that when the motion vectors are uniformly random, the performance bound exhibits a threshold number of frames, above which reasonable performance may be expected with high probability. Furthermore, we highlighted how the CR bound could be used to choose system parameters such as the number of frames for a given downsampling factor to achieve a target performance. Finally, we showed that zero crossings in the imaging system's OTF PSF lead to unstable estimation and that prior information is required to achieve acceptable performance for such situations.

The work presented here suggests several possible research directions. One could imagine using the CR bound for imaging system design under the assumption that super-resolution processing would be included as a postprocessing step. By allowing or even introducing aliasing, one might leverage greater performance out of cheaper imaging system components. Also, it would be interesting to explore novel prior information structures which optimizes the CR bound. For specific imaging applications, novel forms of prior information will not only aid in super-resolution algorithm development, but in imaging system design as well. For instance, if given a set of training images from which to "learn" prior information, one might wonder what is the most beneficial information to glean from the training set for super-resolution. Finally, the analysis presented in this paper focusses on the case of image sequences containing simple translational motion. Future research into super-resolution for more complicated motion models may lead to more broadly applicable results.

## Appendix I
## Fisher Information Matrix for the 2-D Scenario

In this section, we show the necessary derivations for the 2-D version of the FIM used to bound the conditional performance for super-resolution. Recall that the modified forward model in the Fourier domain is

$$\tilde{\mathbf{y}}_k = \tilde{\mathbf{D}}\tilde{\mathbf{F}}(\mathbf{v}_k)\tilde{\mathbf{z}} + \tilde{\mathbf{e}}_k. \tag{52}$$

The vector $\tilde{\mathbf{z}}$ is a $N_H$-dimensional vector with the first $(N_H/2)$ dimensions representing the real components and the the second $(N_H/2)$ dimensions representing the imaginary components. For the 1-D scenario, we used $\omega_i$ to identify the spatial frequency. For the 2-D scenario, we represent the spatial frequencies in the second dimension as $\nu_j$. For the 2-D scenario, all of the matrices have the same structure as the 1-D scenario with the exception that the translation matrix $\tilde{\mathbf{F}}(\mathbf{v}_k)$ is the similar structure as the 1-D case shown in (17), with the exception that the trigonometric terms are now $\cos(v_1\omega_i + v_2\nu_j)$ and $\sin(v_1\omega_i + v_2\nu_j)$.

The log-likelihood function for the observed data is given by

$$\ln p(\{\tilde{\mathbf{y}}_k\} \mid \tilde{\mathbf{z}}, \underline{\mathrm{v}}) = \frac{-1}{2\sigma^2}\sum_{k=0}^{K}(\tilde{\mathbf{y}}_k - \tilde{\mathbf{D}}\tilde{\mathbf{F}}(\mathbf{v}_k)\tilde{\mathbf{z}})^T \times (\tilde{\mathbf{y}}_k - \tilde{\mathbf{D}}\tilde{\mathbf{F}}(\mathbf{v}_k)\tilde{\mathbf{z}}) + \text{const.} \tag{53}$$

Recalling the definition of the FIM in (4), we see that the 2-D FIM is given by (54) and (55), shown at the bottom of the page, where $\tilde{\mathbf{W}}_i$ are the partial derivative operators in the Fourier domain. They are block diagonal matrices of the form

$$\tilde{\mathbf{W}}_1 = \begin{pmatrix} 0 & -\mathrm{diag}(\omega_i) \\ \mathrm{diag}(\omega_i) & 0 \end{pmatrix} \tag{56}$$

$$\tilde{\mathbf{W}}_2 = \begin{pmatrix} 0 & -\mathrm{diag}(\nu_j) \\ \mathrm{diag}(\nu_j) & 0 \end{pmatrix}. \tag{57}$$

Finally, we see that

$$\begin{aligned} &-E_{\tilde{\mathbf{y}}\mid\tilde{\mathbf{z}},\underline{\mathrm{v}}}\left[\frac{\partial^2 l(\{\tilde{\mathbf{y}}_k\}\mid\tilde{\mathbf{z}},\underline{\mathrm{v}})}{\partial\mathbf{v}_k\partial\tilde{\mathbf{z}}}\right] \\ &\quad= \frac{1}{\sigma^2}\left[\tilde{\mathbf{F}}_k^T\tilde{\mathbf{D}}^T\tilde{\mathbf{D}}\tilde{\mathbf{F}}_k\tilde{\mathbf{W}}_1\tilde{\mathbf{z}} \quad \tilde{\mathbf{F}}_k^T\tilde{\mathbf{D}}^T\tilde{\mathbf{D}}\tilde{\mathbf{F}}_k\tilde{\mathbf{W}}_2\tilde{\mathbf{z}}\right] \\ &\quad= \tilde{\mathbf{b}}_k \end{aligned} \tag{58}$$

so that our final FIM is given by $\mathbf{J} = \begin{pmatrix} \tilde{\mathbf{A}} & \tilde{\mathbf{B}} \\ \tilde{\mathbf{B}}^T & \tilde{\mathbf{C}} \end{pmatrix}$ where

$$\tilde{\mathbf{A}} = \sum_{k=0}^{K}\tilde{\mathbf{F}}_k^T\tilde{\mathbf{D}}^T\tilde{\mathbf{D}}\tilde{\mathbf{F}}_k \tag{59}$$

$$\tilde{\mathbf{B}} = [\cdots \quad \tilde{\mathbf{b}}_k \quad \cdots] \tag{60}$$

$$\tilde{\mathbf{C}} = \begin{bmatrix} \tilde{\mathbf{C}}_{11} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \tilde{\mathbf{C}}_{KK} \end{bmatrix}. \tag{61}$$

## Appendix II
## Singular FIM for Translations "On the Grid"

In this section, we show that the 1-D conditional FIM is necessarily singular when the set of translations $\{v_k\}$ are all integers. This corresponds to the canonical example in super-resolution experiments of having the low-resolution frames falling perfectly on the high-resolution "grid" points. In this derivation, it is easier to conceptualize the proof in the spatial domain. Before we begin, we define the quadrants of $\mathbf{J}$ in the spatial domain to be

$$\mathbf{J} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{bmatrix}. \tag{62}$$

---

$$-E_{\tilde{\mathbf{y}}\mid\tilde{\mathbf{z}},\underline{\mathrm{v}}}\left[\frac{\partial^2 l(\{\tilde{\mathbf{y}}_k\}\mid\tilde{\mathbf{z}},\underline{\mathrm{v}}))}{\partial\tilde{\mathbf{z}}^2}\right] = \frac{1}{\sigma^2}\left[\sum_{k=0}^{K}\tilde{\mathbf{F}}_k^T\tilde{\mathbf{D}}^T\tilde{\mathbf{D}}\tilde{\mathbf{F}}_k\right] = \tilde{\mathbf{A}} \tag{54}$$

and

$$-E_{\tilde{\mathbf{y}}\mid\tilde{\mathbf{z}},\underline{\mathrm{v}}}\left[\frac{\partial^2 l(\{\tilde{\mathbf{y}}_k\}\mid\tilde{\mathbf{z}},\underline{\mathrm{v}})}{\partial\mathbf{v}_k^2}\right] = \frac{1}{\sigma^2}\begin{bmatrix} \tilde{\mathbf{z}}^T\tilde{\mathbf{W}}_1^T\tilde{\mathbf{F}}_k^T\tilde{\mathbf{D}}^T\tilde{\mathbf{D}}\tilde{\mathbf{F}}_k\tilde{\mathbf{W}}_1\tilde{\mathbf{z}} & \tilde{\mathbf{z}}^T\tilde{\mathbf{W}}_1^T\tilde{\mathbf{F}}_k^T\tilde{\mathbf{D}}^T\tilde{\mathbf{D}}\tilde{\mathbf{F}}_k\tilde{\mathbf{W}}_2\tilde{\mathbf{z}} \\ \tilde{\mathbf{z}}^T\tilde{\mathbf{W}}_2^T\tilde{\mathbf{F}}_k^T\tilde{\mathbf{D}}^T\tilde{\mathbf{D}}\tilde{\mathbf{F}}_k\tilde{\mathbf{W}}_1\tilde{\mathbf{z}} & \tilde{\mathbf{z}}^T\tilde{\mathbf{W}}_2^T\tilde{\mathbf{F}}_k^T\tilde{\mathbf{D}}^T\tilde{\mathbf{D}}\tilde{\mathbf{F}}_k\tilde{\mathbf{W}}_2\tilde{\mathbf{z}} \end{bmatrix} = \tilde{\mathbf{C}}_{kk} \tag{55}$$

When the translations $v_k$ are multiples of integer sample translations, the matrix $\mathbf{A}^{-1}$ is a diagonal matrix with the terms along the diagonal being $(1/T_i)$ where $T_i$ represents the total number of low-resolution frames with motions $v_k = v_i$ (corresponding to a particular grid location for the high-resolution image). This property has been noted in [14]. We assume that there are $M$ unique translations in the set of all translations and that these translations are all integer offsets of the reference frame (in the high-resolution image coordinates). We use $S_i$ to denote the index set such that $v_k = v_i, \forall k \in S_i$. Without loss of generality, we assume that the unknown translations are ordered such that all $k \in S_i$ are nondecreasing. This ordering induces the structure on $\mathbf{C}$ such that

$$\mathbf{C} = \begin{pmatrix} c_0\mathbf{I}_{T_0} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & c_{M-1}\mathbf{I}_{T_{M-1}} \end{pmatrix} \tag{63}$$

where the subscript $\mathbf{I}_{T_i}$ indicates the dimension of the identity matrix and $c_i = \mathbf{d}^T\mathbf{Q}_i\mathbf{d}$. This ordering also induces structure on the matrix $\mathbf{B}$ such that $\mathbf{b}_k = \mathbf{b}_i, \forall k \in S_i$. Because of the structures of $\mathbf{A}$ and $\mathbf{B}$, we note that $\mathbf{A}^{-1}\mathbf{b}_i = (1/T_i)\mathbf{b}_i$ Thus, we see that $\mathbf{B}^T\mathbf{A}^{-1}\mathbf{B}$ has a block diagonal form

$$\mathbf{B}^T\mathbf{A}^{-1}\mathbf{B} = \begin{pmatrix} \mathbf{M}_0 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \mathbf{M}_{M-1} \end{pmatrix} \tag{64}$$

where

$$\begin{aligned} \mathbf{M}_i &= \left(\frac{1}{T_i}\mathbf{b}_i^T\mathbf{b}_i\right)\mathbf{1}\mathbf{1}^T = \left(\frac{1}{T_i}\mathbf{d}^T\mathbf{Q}_i^T\mathbf{Q}_i\mathbf{d}\right)\mathbf{1}\mathbf{1}^T \\ &= \left(\frac{c_i}{T_i}\right)\mathbf{1}\mathbf{1}^T \end{aligned} \tag{65}$$

where the last equality holds because $\mathbf{Q}$ is a projection operator and, hence, $\mathbf{Q}^T\mathbf{Q} = \mathbf{Q}$.

Thus, we see that the Schur complement Fisher information is given by the block-diagonal matrix

$$\mathbf{S}_{\underline{\mathrm{v}}} = \begin{pmatrix} \mathbf{S}_{A_0} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \mathbf{S}_{A_{M-1}} \end{pmatrix} \tag{66}$$

where

$$\mathbf{S}_{A_i} = \begin{cases} c_i\left[\mathbf{I}_{T_i} - \frac{1}{T_i}\mathbf{1}\mathbf{1}^T\right], & i \neq 0 \\ c_i\left[\mathbf{I}_{T_i} - \frac{1}{T_i+1}\mathbf{1}\mathbf{1}^T\right], & \text{else} \end{cases} \tag{67}$$

Thus, every block along the diagonal other than $i = 0$ is rank deficient by one dimension. This shows that for the very common scenario where all of the the motions are in units of high-resolution pixels (only on the grid), the FIM is rank deficient by $M - 1$.

## Appendix III
## Derivation of the Schur Matrices for $M = 1$

Here, we look at the conditional Schur information complements for the case where there is no aliasing in the model. In this case, we have $\tilde{\mathbf{Q}}_m = \mathbf{I}$ for all $m$. Thus, we see that

$$\tilde{\mathbf{A}} = (K+1)\mathbf{I} \tag{68}$$

$$\tilde{\mathbf{B}} = [\cdots \quad \tilde{\mathbf{d}} \quad \cdots] \tag{69}$$

$$\tilde{\mathbf{C}} = \left(\tilde{\mathbf{d}}^T\tilde{\mathbf{d}}\right)\mathbf{I}. \tag{70}$$

First, we note that the Schur complement of $\tilde{\mathbf{A}}$ is given by

$$\begin{aligned} \tilde{\mathbf{S}}_{\underline{\mathrm{v}}} &= \tilde{\mathbf{C}} - \tilde{\mathbf{B}}^T\tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}} \\ &= (\tilde{\mathbf{d}}^T\tilde{\mathbf{d}})\mathbf{I} - \frac{1}{K+1}\tilde{\mathbf{B}}^T\tilde{\mathbf{B}} \\ &= \tilde{\mathbf{d}}^T\tilde{\mathbf{d}}\left[\mathbf{I} - \frac{1}{K+1}\mathbf{1}\mathbf{1}^T\right]. \end{aligned} \tag{71}$$

Using the matrix inversion lemma [11], we see that

$$\begin{aligned} \left[\mathbf{I} - \frac{1}{K+1}\mathbf{1}\mathbf{1}^T\right]^{-1} &= \mathbf{I} + \mathbf{1}(K+1-\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T \\ &= \mathbf{I} + \mathbf{1}\mathbf{1}^T \end{aligned} \tag{72}$$

where $\mathbf{1}$ represents a vector of all ones. So, the inverse of $\mathbf{S}_{\underline{\mathrm{v}}}$ is given by (using the matrix inversion lemma)

$$\tilde{\mathbf{S}}_{\underline{\mathrm{v}}}^{-1} = \frac{1}{\tilde{\mathbf{d}}^T\tilde{\mathbf{d}}}(\mathbf{I} + \mathbf{1}\mathbf{1}^T). \tag{73}$$

This has the same form as derived previously for looking only at the performance bounds for estimating translational motion [27]. Furthermore, it is interesting to note that, for the case when no aliasing is present, adding additional frames to the problem does not influence the image registration problem. In other words, registration can be done in a pairwise fashion without any loss of information. This is not the case in the presence of aliasing.

To capture the MSE performance in estimating the image terms $\tilde{\mathbf{z}}$, we need only too look at the term

$$\begin{aligned} \mathbf{S}_{\tilde{\mathbf{z}}}^{-1} &= \tilde{\mathbf{A}}^{-1} + \tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}}\mathbf{S}_{\underline{\mathrm{v}}}^{-1}\tilde{\mathbf{B}}^T\tilde{\mathbf{A}}^{-1} \\ &= \frac{1}{K+1}\mathbf{I} + \frac{1}{(K+1)^2}\frac{1}{\tilde{\mathbf{d}}^T\tilde{\mathbf{d}}}(\tilde{\mathbf{B}}\tilde{\mathbf{B}}^T + \tilde{\mathbf{B}}\mathbf{1}\mathbf{1}^T\tilde{\mathbf{B}}^T) \\ &= \frac{1}{K+1}\mathbf{I} + \frac{K+K^2}{(K+1)^2}\frac{\tilde{\mathbf{d}}\tilde{\mathbf{d}}^T}{\tilde{\mathbf{d}}^T\tilde{\mathbf{d}}} \\ &= \frac{1}{K+1}\mathbf{I} + \frac{K}{(K+1)}\frac{\tilde{\mathbf{d}}\tilde{\mathbf{d}}^T}{\tilde{\mathbf{d}}^T\tilde{\mathbf{d}}}. \end{aligned} \tag{74}$$

Finally, we note that, for this simple scenario, the scalar conditional CR bound is given by

$$\begin{aligned} T(\tilde{\mathbf{z}}) &= \left(\frac{\mathrm{tr}(\mathbf{I})}{N_H(K+1)} + \frac{K}{N_H(K+1)}\frac{\tilde{\mathbf{d}}^T\tilde{\mathbf{d}}}{\tilde{\mathbf{d}}^T\tilde{\mathbf{d}}}\right)^{\frac{1}{2}} \\ &= \left(\frac{N_H+K}{N_H(K+1)}\right)^{\frac{1}{2}}. \end{aligned} \tag{75}$$

## Acknowledgment

The authors would like to thank Dr. M. Elad of Technion–Israel Institute of Technology and S. Farsiu of UCSC, as well as the anonymous reviewers for their helpful discussion and comments on the manuscript.

## References

[1] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Advances and challenges in super-resolution," *Int. J. Imag. Syst. Technol.*, vol. 14, no. 2, pp. 47–57, Oct. 2004.

[2] R. Hardie, K. Barnard, and E. Armstrong, "Joint map registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1621–1633, Dec. 1997.

[3] P. Vandewalle, L. Sbaiz, S. Susstrunk, and M. Vetterli, "How to take advantage of aliasing in bandlimited signals," in *Proc. IEEE Conf. Acoustics, Speech, Signal Processing*, May 2004, pp. 948–951.
[4] B. Tom and A. Katsaggelos, "Resolution enhancement of monochrome and color video using motion compensation," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 278–287, Feb. 2001.
[5] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 9, pp. 1167–1183, Sep. 2002.
[6] Z. Lin and H.-Y. Shum, "Fundamental limits on reconstruction-based superresolution algorithms under local translation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 83–97, Jan. 2004.
[7] A. Rajagopalan and P. Kiran, "Motion-free superresolution and the role of relative blur," *J. Opt. Soc. Amer.*, vol. 20, no. 11, pp. 2022–2032, Nov. 2003.
[8] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
[9] H. L. VanTrees, *Detection, Estimation, and Modulation Theory*. New York: Wiley, 1968, pt. I.
[10] R. D. Gill and B. Y. Levit, "Application of the Van Trees inequality: A Bayesian Cramér–Rao bound," *Bernoulli*, no. 1, pp. 59–79, 1995.
[11] F. Graybill, *Matrices with Applications in Statistics*. Belmont, MA: Wadsworth, 1969.
[12] L. Scharf and T. McWhorter, "Geometry of the Crámer–Rao bound," *Signal Process.*, vol. 31, pp. 303–311, 1993.
[13] A. Nehorai and M. Hawkes, "Performance bounds for estimating vector systems," *IEEE Trans. Signal Process.*, vol. 48, no. 6, pp. 1737–1749, Jun. 2000.
[14] M. Elad and Y. Hel-Or, "A fast super-resolution reconstruction algorithm for pure translational motion and common space invariant blur," *IEEE Trans. Image Process.*, vol. 10, no. 8, pp. 1186–1193, Aug. 2001.
[15] B. Porat, *A Course in Digital Signal Processing*. New York: Wiley, 1997.
[16] R. Bracewell, *The Fourier Transform and Its Applications*. New York: McGraw-Hill, 1999.
[17] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Amer. A*, vol. 4, no. 12, pp. 2379–2393, Dec. 1987.
[18] L. Pickup, S. Roberts, and A. Zisserman, "A sampled texture prior for image super-resolution," presented at the Advances in Neural Information Processing Systems 16 Conf., 2004.
[19] W. Freeman, T. Jones, and E. Pasztor, "Example-based super-resolution," *IEEE Comput. Graph. Appl.*, vol. 22, no. 2, pp. 56–65, Feb. 2002.
[20] M. Unser and J. Zerubia, "Generalized sampling: Stability and performance analysis," *IEEE Trans. Signal Process.*, vol. 45, no. 12, pp. 2941–2950, Dec. 1997.
[21] D. Harville, *Matrix Algebra: Exercises and Solutions* New York, 2001.
[22] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super-resolution," *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1327–1344, Oct. 2004.
[23] P. Stoica and T. Marzetta, "Parameter estimation problems with singular information matrices," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 87–90, Jan. 2001.
[24] R. Wagner, R. Baraniuk, and R. Nowak, "Distributed image compression for sensor networks using correspondence analysis and super-resolution," in *Proc. IEEE Int. Conf. Image Processing*, vol. 1, Sep. 2003, pp. 597–600.
[25] A. K. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
[26] H. S. Stone, M. Orchard, E.-C. Chang, and S. Martucci, "A fast direct Fourier-based algorithm for subpixel registration of images," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 10, pp. 2235–2243, Oct. 2001.
[27] D. Robinson and P. Milanfar, "Fundamental performance limits in image registration," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 677–689, Sep. 2004.

**Dirk Robinson** received the B.S. degree in electrical engineering from Calvin College, Grand Rapids, MI, in 1999, and the M.S. degree in computer engineering and the Ph.D. degree in electrical engineering from the University of California, Santa Cruz, in 2001 and 2004, respectively.

He is currently a Research Engineer with Ricoh Innovations, Menlo Park, CA, working in the area of image and signal processing and computational imaging.

**Peyman Milanfar** (SM'98) received the B.S. degree in electrical engineering and mathematics from the University of California, Berkeley, and the S.M., E.E., and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1988, 1990, 1992, and 1993, respectively.

Until 1999, he was a Senior Research Engineer at SRI International, Menlo Park, CA. He is currently Associate Professor of electrical engineering, University of California, Santa Cruz. He was a Consulting Assistant Professor of computer science at Stanford University, Stanford, CA, from 1998 to 2000, where he was also a Visiting Associate Professor from June to December 2002. His technical interests are in statistical signal and image processing and inverse problems.

Dr. Milanfar won a National Science Foundation CAREER award in 2000 and he was Associate Editor for the IEEE SIGNAL PROCESSING LETTERS from 1998 to 2001.