



Published in final edited form as:

IEEE Trans Signal Process. 2013 December 15; 61(24): 6291–6306. doi:10.1109/TSP.2013.2284154.

## Analysis of Point Based Image Registration Errors With Applications in Single Molecule Microscopy

**E. A. K. Cohen [Senior Member, IEEE]** and

Eric Jonsson School of Electrical Engineering and Computer Science, University of Texas at Dallas, Richardson, TX 75083 USA. He is also with the Department of Mathematics, Imperial College London, SW7 2AZ U.K

**R. J. Ober [Senior Member, IEEE]**

Eric Jonsson School of Electrical Engineering and Computer Science, University of Texas at Dallas, Richardson, TX 75083 USA

E. A. K. Cohen: e.cohen@imperial.ac.uk; R. J. Ober: ober@utdallas.edu

### Abstract

We present an asymptotic treatment of errors involved in point-based image registration where control point (CP) localization is subject to heteroscedastic noise; a suitable model for image registration in fluorescence microscopy. Assuming an affine transform, CPs are used to solve a multivariate regression problem. With measurement errors existing for both sets of CPs this is an *errors-in-variable* problem and linear least squares is inappropriate; the correct method being generalized least squares. To allow for point dependent errors the equivalence of a generalized maximum likelihood and heteroscedastic generalized least squares model is achieved allowing previously published asymptotic results to be extended to image registration. For a particularly useful model of heteroscedastic noise where covariance matrices are scalar multiples of a known matrix (including the case where covariance matrices are multiples of the identity) we provide closed form solutions to estimators and derive their distribution. We consider the *target registration error* (TRE) and define a new measure called the *localization registration error* (LRE) believed to be useful, especially in microscopy registration experiments. Assuming Gaussianity of the CP localization errors, it is shown that the asymptotic distribution for the TRE and LRE are themselves Gaussian and the parameterized distributions are derived. Results are successfully applied to registration in single molecule microscopy to derive the key dependence of the TRE and LRE variance on the number of CPs and their associated photon counts. Simulations show asymptotic results are robust for low CP numbers and non-Gaussianity. The method presented here is shown to outperform GLS on real imaging data.

### Keywords

Errors-in-variable; fluorescence microscopy; generalized least squares; image registration

### I. Introduction

Image registration is the process of overlaying two or more images of the same scene [1]. From a theoretical stance, registration is the process of establishing the geometric transformations between two or more data sets such that they can be viewed in a single

coordinate system. These images could arise from different times (multitemporal), different viewpoints (multiview), or different sensors (multimodal).

Broadly speaking, image registration techniques can be divided into two categories. The first is intensity-based registration where gray scale values in both images are correlated to match the images, e.g. [2], [3]. Here we are concerned with the alternative method, feature-based registration, whereby correspondence between the two images is determined through the matching of distinct features common in both images e.g. [4], [5], (although there exist methods that combine both feature and intensity-based approaches to registration e.g. [6]). Specifically we are concerned with the case where the features used for matching are points in the image where pair correspondence is certain. In this case they are known as *control points* (CPs). It is common that these points are created with the use of fiducial markers, e.g. beads in microscopy [7], [8], or infrared emitting diodes in computer aided surgery [9].

We consider an image to capture a subset of the space  $\mathbb{R}^d$ ,  $d = 2$  or  $3$ . Given two image spaces  $\mathcal{A} \subseteq \mathbb{R}^d$  and  $\mathcal{B} \subseteq \mathbb{R}^d$ , say, registration is concerned with estimating the mapping  $T : \mathcal{A} \rightarrow \mathcal{B}$ . It is typical to consider  $T$  to be an *affine* transformation e.g. [10]–[13]. In this circumstance, for  $x \in \mathcal{A}$ ,  $T(x) = Ax + s$  where  $A \in \mathbb{R}^{d \times d}$  is a square invertible matrix and  $s \in \mathbb{R}^d$  is a translation vector. This includes the well studied subclass of *rigid* transformations where the matrix  $A$  is a rotation transformation [11], [12], [14], [15]. Registration involves using the CP locations in  $\mathcal{A}$  and their corresponding mapped positions in  $\mathcal{B}$  to find  $T$ .

In general, due to noisy signals, the location of the CPs in at least one of the images can not be measured exactly and instead are perturbed by random errors. Commonly these error terms are not identically distributed (heteroscedastic) and/or directional (anisotropic). Consequently it is not possible to exactly match the CPs in both images. With this problem in mind, two key questions arise. Firstly; what is the procedure for estimating  $A$  and  $s$  that correctly accounts for the measurement errors in localizing the CPs? Secondly; how accurately can we determine the transformation and hence what errors arise from the registration process? In response to the second of these questions, it has been common in the literature to define the *target registration error* (TRE) as a measure of accuracy for a registration and its distributional properties are of keen interest e.g. [10]–[12], [14]–[16].

One of the most widely researched and applied methods of image registration has been the traditional least squares estimator [13], [14], [17]. Given that the CP locations are precisely known without error in one of the images, and the errors in their localization in the second image are independent and identically distributed (iid) then this provides a proper method of registration. In the case of rigid transformations ( $A$  represents rotation only) [14] provides an approximation to the root mean square of the TRE that has been corroborated with simulated data. This was extended to an approximate distribution of the TRE in [15] and [16]. For rigid transforms when errors are only present in one set of CPs then several papers have attempted to extend distributional results for the TRE to the case where errors are heteroscedastic and anisotropic based on a number of different approaches, including maximum likelihood procedures [12], and a spatial stiffness model [10].

In most registration scenarios measurement errors will exist in both sets of CP locations rendering these methods insufficient. In this circumstance the problem is known as an *errors-in-variables* (EIV) problem and it is well known that the traditional least square method provides inconsistent estimators [18]. If all measurement errors are iid then the *total least squares* (TLS) method (see [19]–[21]) is the correct procedure and [22] provides distributional results for the parameter estimators in the Gaussian case. Under the assumption that measurement errors are iid and white then [23] (corrected by [24]) takes a maximum likelihood (ML) approach to the EIV problem associated with image registration

and Cramér-Rao lower bounds are derived for the variance of parameter estimators. However the reality is iid is a rare luxury and any deviation from this render the TLS and ML methods inconsistent. It is therefore necessary to consider the broader class of model called heteroscedastic EIV (HEIV).

In this paper we will use fluorescence microscopy as a motivating example. Using fiducial markers to perform image registration is an important pre-processing step when correcting for drift between successive frames (multitemporal) e.g. [7], or combining a pair of different colored monochromatic images captured through different sensors (multimodal) e.g. [8], [25]. Localization accuracy depends on the brightness of the light emitting object (see [26]–[30]) and hence each fiducial marker is localized with varying degrees of accuracy. This presents us with a typical HEIV model. It is useful to define a new measure of registration error that we will call the localization registration error (LRE). Recent advances in microscopy have made it possible to detect single molecules in a cellular environment, e.g. [31]–[33]. Localizing a feature (e.g. a single molecule) in  $\mathcal{A}$  typically has its own errors associated with it. The LRE measures the combined effect of this localization error and the registration error to give the localization error of the feature registered in the second image, and is of importance to researchers [8], [25].

There have been recent attempts to tackle the EIV approach to rigid image registration for heteroscedastic errors in [34], [35] with the heteroscedastic EIV (HEIV) algorithm; an iterative procedure that finds an optimal solution to the HEIV model. Numerical Monte Carlo estimates of the TRE for the HEIV algorithm are considered in [11] and compared with a spatial stiffness model approach to the HEIV problem.

In this paper we consider the HEIV model for CP registration, the most general form of the registration problem (under the affine assumption). In Section II we rigorously formulate CP based image registration and formally define the TRE and LRE measures. In Section III, by taking the generalized maximum likelihood approach introduced in [36] as a starting point, we are able to show its equivalence to a heteroscedastic formulation of *generalized least squares* (GLS) and an EIV analogy of the least squares approach, here called *ordinary least squares* (OLS) in keeping with the terminology of [22] for iid EIV. In the case where error covariance matrices for each CP are a scalar multiple of a known matrix (e.g. multiples of the identity), we derive the closed form solution. In Section IV, asymptotic results derived in [36] are used to derive distributions for the registration parameters. In Section V these distributions are used to derive asymptotic distributions for the TRE and LRE. In Section VI we derive the asymptotic second order moments of the TRE and LRE in a microscopy setting, giving neat closed form expressions in terms of photon counts and experimental parameters. We verify these results in Section VII with numerical simulations and show asymptotic results are appropriate for relatively low (realistic) numbers of CPs. The method presented here is shown in Section VIII to outperform the traditional GLS method (that assumes homoscedastic measurement errors) when applied to real fluorescence microscopy imaging data. This paper represents a significant development upon the preliminary results first reported in [37].

A comprehensive list of abbreviations and notations used in this paper can be found in Tables II–V in Appendix A.

## II. Formulating the Problem

Suppose  $K$  CPs are located in at true locations in  $\mathcal{A}_1 \subseteq \mathbb{R}^d$  at true locations  $\{x_{1,k} \in \mathcal{A}_1, k = 1, \dots, K\}$ , and in  $\mathcal{A}_2 \subseteq \mathbb{R}^d$  at true locations, such that  $\{x_{2,k} \in \mathcal{A}_2, k = 1, \dots, K\}$ , such that  $x_{2,k} = T(x_{1,k}) = Ax_{1,k} + s$ ,  $k = 1, \dots, K$ , where  $A \in \mathbb{R}^{d \times d}$  and  $s \in \mathbb{R}^d$ . In reality the positions of the

CPs cannot be known exactly and must instead be measured. Consequently we observe the CP locations as  $\{y_{1,k} \in \mathcal{A}_1, k = 1, \dots, K\}$  and  $\{y_{2,k} \in \mathcal{A}_2, k = 1, \dots, K\}$ , where  $y_{j,k} = x_{j,k} + \epsilon_{j,k}$ ,  $k = 1, \dots, K, j = 1, 2$ . The term  $\epsilon_{j,k} \in \mathbb{R}^d$  is a random variable known as the measurement error. Each measurement error is assumed zero mean and to have individual symmetric positive definite covariance matrix  $\Omega_{j,k}$ . It is assumed that all measurement errors are pairwise independent across the CPs.

We define the  $\mathbb{R}^{d \times K}$  matrices  $X_j \equiv [x_{j,1}, \dots, x_{j,K}]$ ,  $Y_j \equiv [y_{j,1}, \dots, y_{j,K}]$  and  $\mathcal{E}_j^T \equiv [\epsilon_{j,1}, \dots, \epsilon_{j,K}]$ ,  $j = 1, 2$ , and further define the *stacked*  $\mathbb{R}^{2d \times K}$  matrices  $X \equiv [X_1^T, X_2^T]^T$ ,  $Y \equiv [Y_1^T, Y_2^T]^T$  and  $\mathcal{E} \equiv [\mathcal{E}_1^T, \mathcal{E}_2^T]^T$ . With this notation the system of equations can be conveniently represented as the single matrix equation

$$Y = \Lambda X + \alpha \mathbf{1}_K^T + \mathcal{E}, \quad (1)$$

where  $T$  is the matrix transpose,  $\alpha = [\mathbf{0}^T, s^T]^T$ ,  $\Lambda = [I_d, A^T]^T$  and  $\mathbf{1}_K$  is a column vector of length  $K$  with every element taking the value 1. The columns of  $X_1$  are known as the *independent* variables and the columns of  $X_2$  are the *dependent* variables. Models of type (1) where observations of both the dependent and independent variables contain measurement errors are EIV models.

If covariance matrices differ across CPs then they must be known [36]. (In [34] the covariance matrices are unknown, but still require estimation through bootstrapping methods). It is convenient at this point to assume the errors in model (1) follow two possible, but different statistical frameworks. The first shall be called the *second-order framework*  $\mathcal{S}$ .

**Assumption I**—Under  $\mathcal{S}$ , the columns of  $\mathcal{E}$  are independent with  $k$ th column

$\epsilon_k \equiv [\epsilon_{1,k}^T, \epsilon_{2,k}^T]^T$  having mean zero and known symmetric positive definite (SPD) covariance matrix

$$\Omega_k \equiv \text{cov}\{\epsilon_k\} = \begin{bmatrix} \Omega_{1,k} & 0 \\ 0 & \Omega_{2,k} \end{bmatrix}, \quad (2)$$

where  $\text{cov}\{v\}$  denotes the covariance matrix of a vector  $v$ .

The alternative framework shall be called the *distributional framework*  $\mathcal{D}$ .

**Assumption II**—Under  $\mathcal{D}$ , the columns of  $\mathcal{E}$  are independent and of known distribution

with the  $k$ th column  $\epsilon_k \equiv [\epsilon_{1,k}^T, \epsilon_{2,k}^T]^T \stackrel{\text{d}}{=} N_{2d}(\mathbf{0}, \Omega_k)$  where the SPD covariance matrix  $\Omega_k$  is again known and of form (2). (Notation  $\stackrel{\text{d}}{=}$  means ‘equal in distribution’ and  $N_p(\mu, \Sigma)$  denotes the  $p$ -variate normal distribution with mean  $\mu$  and covariance  $\Sigma$ .)

With both  $\mathcal{S}$  and  $\mathcal{D}$ , matrices  $\{\Omega_k, k = 1, \dots, K\}$  are in general not equal. In this circumstance (1) is a HEIV model.

Given measurements  $Y$ , and assuming (1) models the system, the image registration process involves constructing estimators  $\hat{A}$  and  $\hat{s}$  for parameters  $A$  and  $s$ , respectively. The method of estimation will be discussed in Section III.

## A. Image Registration Errors

Suppose we have registered the pair of images with estimators  $\hat{A}$  and  $\hat{s}$ , we need a measure of how successful the registration procedure has been. We begin by defining the commonly used target registration error [10]–[12], [14]–[16]. For anisotropic noise it is important to consider the error as a vector, rather than just its magnitude.

**Definition II.1**—The target registration error (TRE)  $\tau: \mathcal{S}_1 \rightarrow \mathbb{R}^d$  for an arbitrary point  $x_1 \in \mathcal{S}_1$  with corresponding mapped position in  $\mathcal{S}_2$  of  $x_2 = Ax_1 + s$  is given as  $\tau(x_1) \equiv x_2 - (\hat{A}x_1 + \hat{s}) = Ax_1 + s - (\hat{A}x_1 + \hat{s})$ .

We may find it more helpful to consider a related measure. Suppose we are interested in registering a specific feature (e.g. a single molecule) in  $\mathcal{S}_1$  with true position  $x_{1,F} \in \mathcal{S}_1$ , in the second image the true position of this feature is  $x_{2,F} \in \mathcal{S}_2$ , with  $x_{2,F} = Ax_{1,F} + s$ . However, as with the CPs, the position of the feature in  $\mathcal{S}_1$  is actually measured to be at  $y_{1,F} = x_{1,F} + \epsilon_{1,F}$ , where  $\epsilon_{1,F}$  is a measurement error with zero mean and covariance  $\Omega_{1,F}$ . Therefore our estimator for the position of the feature in  $\mathcal{S}_2$  is  $\hat{A}y_{1,F} + \hat{s}$ . A key question is; what is the error associated with localizing the feature in  $\mathcal{S}_2$ ? To quantify this we define a new measure that we will call the *localization registration error*.

**Definition II.2**—For a feature in  $\mathcal{S}_1$  with true and measured locations  $x_{1,F}$  and  $y_{1,F} = x_{1,F} + \epsilon_{1,F}$  respectively, the localization registration error (LRE)  $\ell_F$  is defined as the difference between the true position and the registered position, i.e.  $\ell_F \equiv x_{2,F} - (\hat{A}y_{1,F} + \hat{s}) = Ax_{1,F} + s - (\hat{A}y_{1,F} + \hat{s})$ .

Let us define the difference between the true and estimated values of the transform parameters as  $\Delta A \equiv \hat{A} - A$  and  $\Delta s \equiv \hat{s} - s$ . It can be shown that

$$\tau(x_1) = -\Delta Ax_1 - \Delta s \quad (3)$$

$$\ell_F = -A\epsilon_{1,F} - \Delta A\epsilon_{1,F} - \Delta Ax_{1,F} - \Delta s. \quad (4)$$

We can connect the two as  $\ell_F = -A\epsilon_{1,F} - \Delta A\epsilon_{1,F} + \tau(x_{1,F})$ . When localization of the feature in  $\mathcal{S}_1$  can be achieved exactly, i.e.  $\epsilon_{1,F} = 0$ , then  $\ell_F = \tau(x_{1,F})$ . To derive the distribution of the TRE and LRE, and importantly their respective covariance matrices  $\Omega_\tau$  and  $\Omega_\ell$  it is necessary to know the distributional properties of the terms  $\Delta A$  and  $\Delta s$ .

## III. Parameter Estimation

Parameter estimation for EIV models of type (1) when the columns of the measurement error matrix  $\mathcal{E}$  are iid is well established. The multivariate total least squares (TLS) (see [19]–[21]) or the multivariate generalized least squares (GLS) method [22] solve different but equivalent minimization problems. The parameter estimators that solve the respective minimization problems are exactly known and [22] also derived their asymptotic distributions for Gaussian measurement errors.

The reality is that the covariance matrices  $\{\Omega_k, k = 1, \dots, K\}$  for the stacked measurement errors  $\{\epsilon_k, k = 1, \dots, K\}$  are, in general, not identical and as such the iid assumption is invalid (although pointwise independence is still assumed). Hence, these estimators and the distributional results derived for the iid case are unsuitable for the image registration problem posed here. It is therefore necessary to take a more general approach to formulating the minimization problem and parameter estimation that can take into account

heteroscedastic measurement errors. Definition III.1(i) is given in [36] and Definitions III.1(ii)–(iii) are given here as a generalization to the minimization problems considered in [22].

**Definition III.1**—Consider the observation matrix  $Y$  of measured CP locations that is assumed to arise from model (1) under Assumption I (second-order framework  $\mathcal{S}$ ).

- i. Define likelihood function  $\mathcal{L}(A, s, X_1; Y) \equiv p(Y; A, s, X_1)$ , where

$p(Y; A, s, X_1) = \prod_{k=1}^K p(y_k; A, s, x_{1,k})$  is the joint probability density function (pdf)

for the columns  $y_k \equiv [y_{1,k}^T, y_{2,k}^T]^T$  of observation matrix  $Y = [y_1, \dots, y_K]$ . The ML

estimators  $\hat{s}^{ml}$ ,  $\hat{A}^{ml}$  and  $\hat{X}_1^{ml}$  are defined as  $\arg \max_{s, A, X_1} \mathcal{L}(A, s, X_1; Y)$ .

- ii. For any given  $A$  and  $s$ , the residual vectors  $r_k \equiv y_k - \Lambda x_{1,k} - a$  are zero mean and

have covariance matrix  $\Omega_k$ ,  $k = 1, \dots, K$ . Let  $R(s, A, X_1; Y) = [\Omega_1^{-\frac{1}{2}} r_1, \dots, \Omega_K^{-\frac{1}{2}} r_K]$

and let  $\|\cdot\|_F$  represent the Frobenius matrix norm, the ordinary least squares

(OLS) estimators  $\hat{s}^{ols}$ ,  $\hat{A}^{ols}$  and  $\hat{X}_1^{ols}$  are defined as

$$\arg \min_{s, A, X_1} \|R(s, A, X_1; Y)\|_F = \arg \min_{s, A, X_1} \sum_{k=1}^K r_k^T \Omega_k^{-1} r_k.$$

- iii. For any given  $A$  and  $s$ , the residual vectors  $q_k \equiv y_{2,k} - A y_{1,k} - s$  are zero mean and have covariance matrix  $\Phi_k = U \Omega_k U^T$ ,  $k = 1, \dots, K$ , where  $U = [-A, I_d]$ . Let

$Q(s, A; Y) = [\Phi_1^{-\frac{1}{2}} q_1, \dots, \Phi_K^{-\frac{1}{2}} q_K]$  the generalized least squares (GLS) estimators

$\hat{s}^{gls}$ ,  $\hat{A}^{gls}$  are defined as  $\arg \min_{s, A} \|Q(s, A; Y)\|_F = \arg \min_{s, A} \sum_{k=1}^K q_k^T \Phi_k^{-1} q_k$ .

The maximum likelihood solutions for models of type (1) are discussed in [36]. The number of unknown nuisance parameters  $\{x_{1,k}, k = 1, \dots, K\}$  increases linearly with the number of observations  $\{y_k, k = 1, \dots, K\}$  and are known as incidental parameters [38]. [36] adopts the generalized likelihood approach of [39] to estimate the structural parameters  $A$  and  $s$  in models containing incidental parameters and an iterative procedure for computing the estimators is given. Let  $\text{vec}(Z)$  for a  $r \times s$  matrix  $Z = [z_1, \dots, z_s]$  be defined as the  $rs \times 1$  vector  $[z_1^T, \dots, z_s^T]$  and define  $\beta \equiv [s^T, \text{vec}(A)^T]^T$ .

**Assumption III**—Let  $D$  be a matrix with  $(m, n)$ th element given as  $E\{\beta^2 \mathcal{L}(A, s, X; Y) / \beta_m \beta_n\}$ . All elements of  $D$  exist and  $D$  is non-singular.

Under Assumption III, these estimators of  $A$  and  $s$  are consistent [36], [40]. We now present the key result of this section, the proof of which is found in Appendix B.

**Theorem III.2**—Consider model (1) under Assumption I (second-order framework  $\mathcal{S}$ ), where the random vectors  $\{\epsilon_k, k = 1, \dots, K\}$  are pairwise independent and vector  $\epsilon_k$  has covariance  $\Omega_k$ ,  $k = 1, \dots, K$ , then the solutions to the OLS and GLS minimization problems (see Definitions III.1(ii) and III.1(iii)) are identical to the ML estimators for the likelihood function under Assumption II (distributional framework  $\mathcal{D}$ ) (i.e.  $\hat{A}^{ml} = \hat{A}^{ols} = \hat{A}^{gls}$ ,  $\hat{s}^{ml} = \hat{s}^{ols} = \hat{s}^{gls}$  and  $\hat{X}_1^{ml} = \hat{X}_1^{ols}$ ), where  $\epsilon_k \stackrel{d}{=} N_{2d}(\mathbf{0}, \Omega_k)$ .

We note the OLS and GLS estimators do not depend on having Gaussian distributed measurement errors.

## A. Weighted Covariance Generalized Least Squares

While iterative procedures are required to compute the ML/OLS/GLS estimators—see Definitions III.1(i),(ii) and (iii)—in the general case of heteroscedastic noise it is possible to derive an exact closed form expression for the ML/OLS/GLS estimators when we consider the following special case, which we term *weighted covariance*.

**Assumption IV**—For all  $k = 1, \dots, K$ , the SPD covariance matrix  $\Omega_k$  of  $\epsilon_k$ —the  $k$ th column of the measurement error matrix  $\mathcal{E}$ —is given by  $\Omega_k = \eta_k \Omega_0$  where  $\Omega_0$  is a known SPD matrix and  $\eta_k \in \mathbb{R}^+$  is known.

Assumption IV will be shown to be suitable in fluorescence microscopy image registration—see Section VI-A—and from a theoretical stance includes the important case of

$\Omega_k = \sigma_k^2 I_{2d}$ . We introduce a further set of assumptions that are necessary for consistency of the estimators presented in Theorem III.3.

### Assumption V

- i. Define the scalar  $\gamma^{(K)} \equiv \left(\frac{1}{K}\right) \sum_{k=1}^K \eta_k^{-1}$ , vector  $\bar{y}^{(K)} \equiv \left(\frac{1}{K}\right) \sum_{k=1}^K \eta_k^{-1} y_k$  and matrix  $W^{(K)} \equiv \left(\frac{1}{K}\right) \sum_{k=1}^K \eta_k^{-1} (y_k - \gamma^{(K)-1} \bar{y}^{(K)}) (y_k - \gamma^{(K)-1} \bar{y}^{(K)})^T$ , then we assume  $\lim_{K \rightarrow \infty} \gamma^{(K)}$  exists, and there exists  $\bar{y}$  and  $W$  such that  $\bar{y} = \lim_{K \rightarrow \infty} \bar{y}^{(K)}$  and  $W = \lim_{K \rightarrow \infty} W^{(K)}$  with probability one (wp1).
- ii. Define vectors  $\bar{x}_1^{(K)} \equiv \left(\frac{1}{K}\right) \sum_{k=1}^K \eta_k^{-1} x_{1,k}$ , matrix  $\Xi^{(K)} \equiv \left(\frac{1}{K}\right) \sum_{k=1}^K \eta_k^{-1} x_{1,k} x_{1,k}^T$  and matrix  $\Psi^{(K)} \equiv \Xi^{(K)} - \gamma^{(K)-1} \bar{x}_1^{(K)} \bar{x}_1^{(K)T}$ , then we assume there exists  $\Psi$  such that  $\Psi = \lim_{K \rightarrow \infty} \Psi^{(K)}$  wp1.

The proof of the following theorem is found in Appendix C.

**Theorem III.3**—Consider the multivariate EIV model (1) under Assumption I (second-order framework  $\mathcal{S}$ ) and Assumption IV (weighted covariance). Define

$\gamma^{(K)} \equiv \left(\frac{1}{K}\right) \sum_{k=1}^K \eta_k^{-1}$ , vector  $\bar{y}^{(K)} \equiv \left(\frac{1}{K}\right) \sum_{k=1}^K \eta_k^{-1} y_k$  and matrix  $W^{(K)} \equiv \left(\frac{1}{K}\right) \sum_{k=1}^K \eta_k^{-1} (y_k - \gamma^{(K)-1} \bar{y}^{(K)}) (y_k - \gamma^{(K)-1} \bar{y}^{(K)})^T$ . The eigendecomposition of  $W^{(K)} \Omega_0^{-1}$  is represented as  $W^{(K)} \Omega_0^{-1} = G E G^{-1}$  with  $E = \text{diag}\{e_1, e_2, \dots, e_{2d}\}$  where  $e_1 \dots e_{2d}$  are the ordered eigenvalues of  $W^{(K)} \Omega_0^{-1}$  and the columns of  $G$  are the

corresponding eigenvectors. Making the partition  $G = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix}$ , and assuming  $G_{11}^{-1}$  exists, the ML/OLS/GLS estimators  $A$  of  $s$  and are given as

$$\hat{A} = G_{21} G_{11}^{-1}, \hat{s} = \gamma^{(K)-1} \hat{U} \bar{y}^{(K)}, \quad (5)$$

where  $\hat{U} = [-\hat{A}, I_d]$ . Furthermore, provided Assumption V additionally holds then  $\hat{A}$  and  $\hat{s}$  are consistent estimators of  $A$  and  $s$ , respectively.

## IV. Asymptotic Distributions for Parameter Estimators

For the most general case where the covariance matrices  $\{\Omega_k, k = 1, \dots, K\}$  are unrelated and unequal SPD matrices, [36] also derived the asymptotic distributions for the ML estimators of  $s$  and  $A$  under distributional framework  $\mathcal{D}$ . These estimators form the elements of the estimator  $\hat{\beta}$  for the parameter vector  $\beta \equiv [s^T, \text{vec}(A)^T]^T$ . Provided Assumption II (statistical framework  $\mathcal{D}$ ) holds, i.e. measurement errors are Gaussian, the result will also be appropriate for the OLS and GLS estimators as given in Definitions III.1(ii) and III.1(iii). We introduce  $\xrightarrow{K}$  to denote asymptotically equal in distribution with respect to  $K$ . The following is from [36].

**Theorem IV.1**—Consider the multivariate EIV model (1) under Assumption II (distributional framework  $\mathcal{D}$ ), then given  $\hat{\beta} \equiv [\hat{s}^T, \text{vec}(\hat{A})^T]^T$  is a consistent estimator of  $\beta \equiv [s^T, \text{vec}(A)^T]^T$ , defining  $\Delta\beta \equiv \hat{\beta} - \beta$  we have  $K^{\frac{1}{2}}\Delta\beta \xrightarrow{K} N_{(d^2+d)}(\mathbf{0}, B)$ , where  $B = \lim_{K \rightarrow \infty} B^{(K)}$ , with  $B^{(K)} = H^{(K)-1} P^{(K)} (H^{(K)-1})^T$ ,

$$H^{(K)} = K^{-1} \sum_{k=1}^K \begin{bmatrix} -\Phi_k^{-1} & -\left(x_{1,k}^T \otimes \Phi_k^{-1}\right) \\ -\left(x_{1,k} \otimes \Phi_k^{-1}\right) & -\left(x_{1,k} x_{1,k}^T \otimes \Phi_k^{-1}\right) \end{bmatrix} \quad (6)$$

and

$$P^{(K)} = K^{-1} \sum_{k=1}^K \begin{bmatrix} \Phi_k^{-1} & -\left(x_{1,k}^T \otimes \Phi_k^{-1}\right) \\ -\left(x_{1,k} \otimes \Phi_k^{-1}\right) & L \otimes \Phi_k^{-1} \end{bmatrix}, \quad (7)$$

where  $\Phi_k = U\Omega_k U^T$ ,  $U = [-A, I_d]$ ,  $L = \left(\Lambda^T \Omega_k^{-1} \Lambda\right)^{-1} + x_{1,k} x_{1,k}^T$  and  $\otimes$  denotes the Kronecker product. Given  $\hat{\beta}$  is a consistent estimator of  $\beta$ , then  $H$  and  $P$  are consistently estimated by  $\hat{H}^{(K)}$  and  $\hat{P}^{(K)}$  respectively, calculated by substituting  $X_1 = [x_{1,1}, \dots, x_{1,K}]$  for  $X_1$ , and then  $\hat{A}$  and  $\hat{s}$  into (6) and (7).<sup>1</sup>

Using this result we are able to present the following corollary for the asymptotic distributions of the estimators in the special case of the weighted covariance property—see Section III-A. We introduce the notation  $[M]_{mn}$  for the  $(m, n)$ th element of a matrix  $M$ , and specifically use  $\Delta a_{mn}$  as short-hand for  $[\Delta A]_{mn}$ . The proof of the following is found in Appendix D.

**Corollary IV.2**—Consider the multivariate EIV model (1) under Assumption II (distributional framework  $\mathcal{D}$ ) and Assumption IV (weighted covariance). Define vector

$$\begin{aligned} \bar{x}_1^{(K)} &\equiv \left(\frac{1}{K}\right) \sum_{k=1}^K \eta_k^{-1} x_{1,k}, \text{ matrix } \Xi^{(K)} \equiv \left(\frac{1}{K}\right) \sum_{k=1}^K \eta_k^{-1} x_{1,k} x_{1,k}^T, \text{ matrix} \\ \Psi^{(K)} &\equiv \Xi^{(K)} - \gamma^{(K)-1} \bar{x}_1^{(K)} \bar{x}_1^{(K)T}, \text{ matrix } \Phi_0 = U\Omega_0 U^T \text{ (where } U = [-A, I_d]), \text{ and matrix} \\ \Theta^{(K)} &= \Psi^{(K)-1} (\Lambda^T \Omega_0^{-1} \Lambda)^{-1} \Psi^{(K)-1} + \Psi^{(K)-1}. \end{aligned}$$

Let  $\hat{\beta} \equiv [\hat{s}^T, \text{vec}(\hat{A})^T]^T$  be constructed from

<sup>1</sup>A minus sign, believed to be incorrectly missing in [36], has been added to the leading diagonal term in (6). This is required for consistency with the well established iid results found in [22] and [41].



the estimators in (5), then under Assumption V (consistency)  $K^{\frac{1}{2}}\Delta\beta\overset{\text{K}}{\underset{\text{K}}{\rightarrow}}N(\mathbf{0}, B)$ , where  $B = \lim_{K \rightarrow \infty} B^{(K)}$ ,  $B^{(K)} \equiv (H_0^{(K)-1} P_0^{(K)} H_0^{(K)-1}) \otimes \Phi_0$  where

$$H_0^{(K)} = \begin{bmatrix} -\gamma^{(K)} & -\bar{x}_1^{(K)T} \\ -\bar{x}_1^{(K)} & -\Xi^{(K)} \end{bmatrix} \quad (8)$$

$$P_0^{(K)} = \begin{bmatrix} \gamma^{(K)} & -\bar{x}_1^{(K)T} \\ -\bar{x}_1 & \Xi^{(K)} + (\Lambda^T \Omega_0^{-1} \Lambda)^{-1} \end{bmatrix}. \quad (9)$$

This gives the following key asymptotic identities as  $K \rightarrow \infty$

$$\begin{aligned} \text{cov} \left\{ K^{\frac{1}{2}} \Delta s_m, K^{\frac{1}{2}} \Delta s_n \right\} &\rightarrow (\gamma^{-1} + \gamma^{-2} \theta) [\Phi_0]_{mn} \\ \text{cov} \left\{ K^{\frac{1}{2}} \Delta s_l, K^{\frac{1}{2}} \Delta a_{mn} \right\} &\rightarrow [\gamma^{-1} \Theta \bar{x}_1]_n [\Phi_0]_{ml} \\ \text{cov} \left\{ K^{\frac{1}{2}} \Delta a_{mn}, K^{\frac{1}{2}} \Delta a_{m'n'} \right\} &\rightarrow [\Theta]_{mm'} [\Phi_0]_{nn'}. \end{aligned}$$

where we define  $\theta \equiv \bar{x}_1^T \Theta \bar{x}_1$  and the limits  $\gamma = \lim_{K \rightarrow \infty} \gamma^{(K)}$  and  $\Theta = \lim_{K \rightarrow \infty} \Theta^{(K)}$  (which both exist under Assumption V).

We note that in the iid case where  $\Omega_k = \sigma^2 I_{2d}$  for all  $k = 1, \dots, K$  then the parameter estimation procedures (as outlined in Theorem III.3) and the asymptotic results (as outlined in Corollary IV.2) reduce to the results presented in [22] and thus forms a natural extension to well established GLS results.

## V. A symptotic Distributions for TRE and LRE

Given the distributional results of Theorem IV.1 and Corollary IV.2 it is now possible to derive asymptotic distributions for the TRE and LRE—see Definitions II.1 and II.2, respectively. The proof of the following is omitted but follows directly from the linear combination of Gaussian random variables being itself Gaussian. The covariance matrix follows directly from (3), Theorem IV.1 and Corollary IV.2.

**Corollary V.1**—Under model (1), Assumption II (distributional framework  $\mathcal{D}$ ) and

Assumption III (consistency), for a point  $x_{1,F} \in \mathcal{X}_1$  then  $K^{\frac{1}{2}} \tau(x_{1,F}) \overset{\text{K}}{\underset{\text{K}}{\rightarrow}} N_d(\mathbf{0}, T)$  with  $(m, n)$ th element given as

$$[T]_{mn} = [B]_{mn} + \sum_{j=1}^d ([x_{1,F}]_j [B]_{(j-1)d+m,n} + [x_{1,F}]_j [B]_{(j-1)d+n,m}) + \sum_{j=1}^d \sum_{j'=1}^d [B]_{(j-1)d+m,(j'-1)d+n} [x_{1,F}]_j [x_{1,F}]_{j'}, \quad (10)$$

with  $B$  as defined in Theorem IV.1. Further to this, under Assumption IV (weighted covariance) and Assumption V (consistency), where  $\Theta$  and  $\Phi_0$  are as defined in Corollary IV.2

$$\begin{aligned}
[T]_{mn} &= (\gamma^{-1} + \gamma^{-1} \bar{x}_1 \Theta \bar{x}_1^T) [\Phi_0]_{mn} \\
&+ \sum_{j=1}^d [x_{1,F}]_j [\gamma^{-1} \Theta \bar{x}_1]_j [\Phi_0]_{mn} \\
&+ \sum_{j=1}^d [x_{1,F}]_j [\gamma^{-1} \Theta \bar{x}_1]_j [\Phi_0]_{nm} \\
&+ \sum_{j=1}^d \sum_{j'=1}^d [\Theta]_{mn} [\Phi_0]_{jj'} [x_{1,F}]_j [x_{1,F}]_{j'}.
\end{aligned} \tag{11}$$

To consider distributional results for the LRE, it is necessary to decompose  $\ell_F$  into two parts,  $\ell_F = -A\varepsilon_{1,F} + z_F$  where  $z_F = -\Delta A\varepsilon_{1,F} - \Delta s - \Delta A x_{1,F}$ . The term  $-A\varepsilon_{1,F}$  is a random variable with no dependency on the image registration process, being independent of  $\Delta A$  and  $\Delta s$  (and hence  $K$ , CP locations and CP measurement errors). By contrast, each term in  $z_F$  is dependent on the image registration process, being a function of  $\Delta A$  and  $\Delta s$ . It is important to note  $z_F$  and  $-A\varepsilon_{1,F}$  are independent. We can now provide the following asymptotic result for  $z_F$ , the proof of which is omitted but follows directly from the linear combination of Gaussian random variables being itself Gaussian. The covariance matrix follows directly from (4), Theorem IV.1 and Corollary IV.2.

**Corollary V.2**—Under model (1), Assumption II (distributional framework  $\mathcal{D}$ ) and Assumption III (consistency) provided  $\varepsilon_{1,F} \stackrel{d}{=} N_d(\mathbf{0}, \Omega_{1,F})$  where  $\varepsilon_{1,F}$  is the measurement error associated with localizing a feature in  $\mathcal{A}_1$  with true location  $x_{1,F}$ , then  $K^{\frac{1}{2}} z_F \stackrel{d}{=} N_d(\mathbf{0}, Z)$  with  $(m, n)$ th element of  $Z$  given as

$$[Z]_{mn} = [T]_{mn} + \sum_{j=1}^d \sum_{j'=1}^d [\Omega_{1,F}]_{jj'} [B]_{(j-1)d+m, (j'-1)d+n}. \tag{12}$$

Further to this, under Assumption IV (weighted covariance) and Assumption V (consistency),

$$[Z]_{mn} = [T]_{mn} + \sum_{j=1}^d \sum_{j'=1}^d \Theta_{mn} [\Phi_0]_{jj'} [\Omega_{1,F}]_{jj'}. \tag{13}$$

While the asymptotic distributions presented here are mathematically elegant, experimenters are keen to know the covariance matrices  $\Omega_\tau$  and  $\Omega_\ell$  of the TRE  $\tau$  and LRE  $\ell_F$ , respectively, in their registration procedure. It is therefore necessary to have results for finite  $K$ . Consider model (1) under Assumptions II (distributional framework  $\mathcal{D}$ ) and III (consistency), or Assumptions II, IV (weighted covariance) and V (consistency). Let  $x_{1,F} \in \mathcal{A}_1$  be the true position of a feature that is localized at  $y_{1,F} = x_{1,F} + \varepsilon_{1,F}$ , where  $\varepsilon_{1,F} \stackrel{d}{=} N_{2d}(\mathbf{0}, \Omega_{1,F})$ . For “large  $K$ ” we assume the asymptotics have been approximately met and hence from Corollaries V.1 and V.2, together with the independence of  $z_F$  and  $-A\varepsilon_{1,F}$ , the TRE  $\tau(x_{1,F})$  and LRE  $\ell_F$  are both approximately  $d$ -dimensional normally distributed with zero mean and have respective approximate covariance matrices

$$\Omega_\tau \approx K^{-1}T, \quad (14)$$

$$\Omega_\ell \approx A\Omega_{1,F}A^T + K^{-1}Z. \quad (15)$$

These approximations will be used in Sections VI and VII to investigate in more detail registration problems in single molecule microscopy.

## VI. Image Registration Error Analysis for Single Molecule Microscopy

We apply the theory presented thus far in this paper to the important problem of assessing localization errors due to the registration process in a fluorescence microscopy setup. Registration is a common pre-processing step in microscopy experiments. Fiducial markers are used as CPs, these are bright, light emitting objects (e.g. fluorescent beads [8], [42]) that require estimation of their location. One example of a fluorescence microscopy experimental setup is to register a pair of different colored monochromatic images, for example to see if two different protein molecules colocalize e.g. [8], [25], [43], [44]. This is a multimodal registration problem. Fiducial markers are also used for drift correction between successive image frames [7]. This is a multitemporal registration problem.

As will now be discussed, the covariance matrix associated with localizing each CP/fiducial marker in each image is dependent on the number of photons associated with it that are detected at the sensor, and therefore presents us with a HEIV problem of type (1).

### A. Measurement Errors

In [26] and [30] are lower bound expressions for the covariance matrix of the error in localizing an isolated point source emitting photons as an inhomogeneous Poisson process in the presence of background and readout noise, which in turn is shown in [26] to be a reasonable estimate for the true covariance matrix. These general expressions can be used with the estimation procedure of [36]—see Section III—for parameter estimation and TRE and LRE second-order moments can be computed with the generalized expressions of (10) and (12), respectively.

When imaging in the absence of background noise and readout noise the covariance matrix for the error in localizing a point source in the *object space* is given as  $N^{-1}J$ , where  $N$  is the number of photons collected from the point source at the detector, and  $J$  is a SPD matrix that can be computed from experimental parameters including photon wavelength, numerical aperture and the point spread function of the optical system. For a non-pixelated detector  $J$  is diagonal, with pixels introducing off-diagonal terms. In the image space this covariance matrix becomes  $N^{-1}\mathcal{M}^2J$ , where  $\mathcal{M}$  is the known *system magnification* between the object space and the image space (a distance  $r$  in the object space is measured as  $\mathcal{M}r$  in the image space). Further to this, it is shown in [45] that even in the presence of typical levels of background and readout noise the covariance matrix  $N^{-1}\mathcal{M}^2J$  is a suitable approximation to the covariance matrix for errors in localizing a bright (high signal to noise ratio) point source. Conventional fiducial markers used in fluorescence microscopy are typically bright and hence for the purposes of this paper we assume that they can be treated in this way.

We assume the image registration formulation of Section II and model (1) with the use of  $K$  fiducial markers for the CPs. The matrix  $J$  and system magnification  $\mathcal{M}$  are specific to the image and hence labeled  $J_j$  and  $\mathcal{M}_j$ , respectively,  $j = 1, 2$ . Suppose  $N_{j,k}$  photons are detected at the detector for fiducial marker  $k$  in  $\mathcal{S}_j$ ,  $k = 1, \dots, K$ ,  $j = 1, 2$ . The measurement errors  $\epsilon_{j,k}$ ,  $k = 1, \dots, K$ ,  $j = 1, 2$ , are therefore assumed to have covariance  $\Omega_{j,k}$  of the form  $\Omega_{j,k} = (1/$

$N_{j,k})\Omega_{j,0}$ , where  $\Omega_{j,0} \equiv \mathcal{M}_j^2 J_j$  is a SPD matrix and universal for all CPs in  $\mathcal{S}_j$ . This gives the covariance matrix of  $\epsilon_k \equiv [\epsilon_{1,k}^T, \epsilon_{2,k}^T]^T$  as the block diagonal matrix  $\Omega_k = \text{diag}\{(1/N_{1,k})\Omega_{1,0}, (1/N_{2,k})\Omega_{2,0}\}$ .

Consider the two common image registration scenarios described at the beginning of this section. The first is in registering two monochromatic images captured at the same time with two different sensors (multimodal). While the photon count associated with a single fiducial marker at two different wavelengths (i.e. in separate monochromatic images) will be different, it is reasonable to assume that there is a linear relation between the brightness of the marker in each image i.e. a marker that is bright in  $\mathcal{S}_1$  is also bright in  $\mathcal{S}_2$ . Mathematically we say  $N_{2,k} = cN_{1,k}$  for all  $k = 1, \dots, K$ , where  $c > 0$  is a constant of proportionality, suitable for all  $k = 1, \dots, K$ . The second scenario is performing drift correction by registering two images taken by the same sensor at different times (multitemporal). In this case, provided the brightness of the marker remains constant in the time between captures, then we assume  $N_{1,k} \approx N_{2,k}$  (which is mathematically equivalent to the multimodal scenario with  $c = 1$ ). The constant  $c$  need not be known to derive expressions for  $\Omega_\tau$  and  $\Omega_\ell$  we just assume it exists.

With this assumption we have the situation where the covariance matrices of the measurement errors have the weighted co-variance property, i.e. they are scalar multiples of

$$\Omega_0 = \begin{bmatrix} \Omega_{1,0} & 0 \\ 0 & c^{-1}\Omega_{2,0} \end{bmatrix} \quad (16)$$

with  $\eta_k = 1/N_{1,k}$  providing the scaling factor. This gives

$\gamma^{(K)} \equiv \left(\frac{1}{K}\right) \sum_{k=1}^K \eta_k^{-1} = \left(\frac{1}{K}\right) \sum_{k=1}^K N_{1,k} = \overline{N}_1^{(K)}$ , the mean photon count for the  $K$  CPs in  $\mathcal{S}_1$ . Fig. 1 summarizes the key steps involved in registering a pair of fluorescence microscopy images.

## B. Microscopy LRE

We now consider the expressions for the TRE and LRE covariance matrices,  $\Omega_\tau$  and  $\Omega_\ell$ , respectively. Consider localizing a single molecule in  $\mathcal{S}_1$  at point  $y_{1,F}$ , where  $y_{1,F} = x_{1,F} + \epsilon_{1,F}$  with the measurement error having covariance  $\Omega_{1,F}$  estimable from [26], and  $x_{1,F}$  being the true positional vector. We estimate its location (the registered position) in  $\mathcal{S}_2$  as  $\hat{A}y_{1,F} + \hat{s}$ .

We begin by considering the following 2D model.

**Assumption (i)**—We model the CP measurement errors  $\epsilon_k \stackrel{d}{=} N_4(0, \Omega_k)$  where  $\Omega_k$  is given in (2) with  $\Omega_{j,k} = \sigma_{j,k}^2 I_2 = (1/N_{j,k}) \cdot \mathcal{M}_j^2 \zeta_j I_2$  where  $N_{j,k}$  is the photon count at the detector associated with CP  $k$  in  $\mathcal{S}_j$ ,  $k = 1, \dots, K$ ,  $j = 1, 2$ .  $\mathcal{M}_j$  is the known system magnification associated with  $\mathcal{S}_j$ .  $\zeta_j$  is a known function of the point spread function, photon wavelength and numerical aperture. In multimodal registration  $\mathcal{M}_j$  and  $\zeta_j$  will be different for each image, while in multitemporal registration they will be identical for both images.  $\Omega_0$  is of form (16) where  $\Omega_{j,0} = \mathcal{M}_j^2 \zeta_j I_2$ .

**Assumption (ii)**—Consider the CP true positions  $\{x_{1,k}, k = 1, \dots, K\}$  to be  $K$  realizations of a random variable  $x \in \mathbb{R}^2$  with mean zero and covariance  $\kappa^2 I_2$ , and let associated photon counts be non-zero, finite and independent of CP positions.

**Assumption (iii)**—The affine transformation parameter  $A = SR$  represents a scaling  $S = \zeta I_2$ ,  $\zeta \in \mathbb{R}^+$ , combined with a unitary rotation or reflection (or a combination of both)  $R$ , i.e.  $R^T R = RR^T = I_2$ .

We will make use of the following Lemma, the proof of which is found in Appendix E.

**Lemma VI.1**—Let CP positions  $\{x_{1,k}, k = 1, \dots, K\}$  be  $K$  realizations of a random variable  $x \in \mathbb{R}^d$  with finite mean  $\mu$  and SPD covariance  $\Pi$ , and let weights  $\{\eta_k, k = 1, \dots, K\}$  be  $K$  realizations of a random variable  $\mathcal{N}$  with finite non-zero mean and finite variance

independent of CP position. Let  $\bar{x}_1^{(K)} \equiv \left(\frac{1}{K}\right) \sum_{k=1}^K \eta_k^{-1} x_{1,k}$  and  $\Psi^{(K)} \equiv \Xi^{(K)} - \gamma^{(K)-1} \bar{x}_1^{(K)} \bar{x}_1^{(K)T}$  where  $\gamma^{(K)} \equiv \left(\frac{1}{K}\right) \sum_{k=1}^K \eta_k^{-1}$  and  $\Xi^{(K)} \equiv \left(\frac{1}{K}\right) \sum_{k=1}^K \eta_k^{-1} x_{1,k} x_{1,k}^T$ , then  $\bar{x}_1 \equiv \lim_{K \rightarrow \infty} \bar{x}_1^{(K)} = E\{\mathcal{N}^{-1}\}\mu$  and  $\Psi \equiv \lim_{K \rightarrow \infty} \Psi^{(K)} = E\{\mathcal{N}^{-1}\}\Pi$ , each wp1.

From Assumption (ii) and Lemma VI.1  $x_1 = 0$  and  $N_1^{-1} \kappa^2 I_2$  will provide our estimate of  $\Psi$ ,

where  $\bar{N}_1 = \left(\frac{1}{K}\right) \sum_{k=1}^K N_{1,k} = \gamma^{(K)}$ . We introduce  $\bar{\sigma}_j^2 \equiv \mathcal{M}_j^2 \zeta_j / \bar{N}_j$ . Using these expressions for  $x_1$  and  $\Psi$  and with the covariance matrix in Assumption (i) and the form of the transformation parameters in Assumption (iii), then from the definitions in Corollary IV. 2 we have the identities  $\Phi_0 = (\zeta^2 \zeta_1 + c^{-1} \zeta_2) I_2$  and  $\Theta = \bar{N}_1^{-1} \kappa^{-4} (\bar{\sigma}_1^{-2} + \zeta^2 \bar{\sigma}_2^{-2})^{-1} I_2 + \bar{N}_1^{-1} \kappa^{-2} I_2$ . Substituting these expressions into (11) and (13) gives

$$[T]_{mn} = (\zeta^2 \bar{\sigma}_1^2 + \bar{\sigma}_2^2) \left( 1 + \frac{r_F^2}{\kappa^2} \left( 1 + \frac{1}{\frac{\kappa^2}{\bar{\sigma}_1^2} + \frac{\zeta^2 \kappa^2}{\bar{\sigma}_2^2}} \right) \right) \delta_{mn},$$

$$[z]_{mn} = [T]_{mn} + (\zeta^2 \bar{\sigma}_1^2 + \bar{\sigma}_2^2) \frac{2\sigma_F^2}{\kappa^2} \left( 1 + \frac{1}{\frac{\kappa^2}{\bar{\sigma}_1^2} + \frac{\zeta^2 \kappa^2}{\bar{\sigma}_2^2}} \right) \delta_{mn}$$

where  $r_F \equiv \|x_{1,F}\|$  is the radial distance of the feature/single molecule from the origin. Given the spread of the CPs (represented by  $\kappa$ ) is much greater than the localization accuracy of the individual CPs (represented by  $\sigma_1$  and  $\sigma_2$ ) and the localization accuracy of

the feature (represented by  $\sigma_F$ )—as is typical in microscopy—then the term  $1 / \left( \frac{\kappa^2}{\bar{\sigma}_1^2} + \frac{\zeta^2 \kappa^2}{\bar{\sigma}_2^2} \right)$  and  $\sigma_F^2 / \kappa^2$  (both in the order of  $10^{-6}$  for a typical microscopy experiment) can be considered negligible. From Corollaries V.1 and V.2 and assuming approximations (14) and (15) we state the following key result:

**Proposition VI.2**—For large  $K$ , under Assumptions (i)–(iii) the  $(m, n)$ th element of the covariance matrix of the TRE can be approximated as

$$[\Omega_\tau]_{mn} \approx K^{-1} \left( \zeta^2 \frac{\mathcal{M}_1^2 \zeta_1}{N_1} + \frac{\mathcal{M}_2^2 \zeta_2}{N_2} \right) \left( 1 + \left( \frac{r_F}{\kappa} \right)^2 \right) \delta_{mn}. \quad (17)$$

The  $(m, n)$  th element of the covariance matrix of the LRE can be approximated as

$$[\Omega_\ell]_{mn} \approx [A\Omega_{1,F}A^T]_{mn} + K^{-1} \left( \zeta^2 \frac{\mathcal{M}_1^2 \zeta_1}{N_1} + \frac{\mathcal{M}_2^2 \zeta_2}{N_2} \right) \left( 1 + \left( \frac{r_F}{\kappa} \right)^2 \right) \delta_{mn}. \quad (18)$$

If the covariance  $\Omega_{1,F}$  is itself representable as  $\sigma_F^2 I_2 = (\mathcal{M}_1^2 \zeta_1 / N_{1,F}) I_2$ , where  $N_{1,F}$  is the photon count associated with the feature (molecule) imaged in  $\mathcal{A}_1$ , then in (18) we replace

$[A\Omega_{1,F}A^T]_{mn}$  with  $\zeta^2 \frac{\mathcal{M}_1^2 \zeta_1}{N_F} \delta_{mn}$ . The covariance matrices  $\Omega_\tau$  and  $\Omega_\ell$  are both given with respect to image space  $\mathcal{S}_2$ . To express these matrices with respect to the object space coordinates we use  $\mathcal{M}_2^{-2} \Omega_\tau$  and  $\mathcal{M}_2^{-2} \Omega_\ell$ .

Consider (17). We immediately notice that the TRE has an inverse dependence the number of CPs (denoted by  $K$ ). Given extra CPs have similar localization errors to those already deployed, then increasing the number of CPs will improve the TRE covariance and it vanishes to zero as the number of CPs tends to infinity. It will be the case that for relatively small numbers of CPs the TRE improvements are significant by attempting to introduce more of them into the registration process. However, if we already have a large number of CPs then there is no significant gain in registration performance by small increases in their numbers. Specifically to microscopy, the TRE covariance will also vanish as the mean photon counts tend to infinity. Therefore extending exposure time will improve registration. Diagonal terms of the TRE covariance have dependency on CP measurement errors in both images, while off diagonal terms of the TRE covariance depend only on measurement errors in  $\mathcal{S}_2$ . In relation to the parameters  $A$  and  $s$ , the LRE and TRE covariance matrices are independent of rotation and translation and exhibit dependence only on scaling factor  $\zeta$ . The

diagonal terms of the  $\Omega_\ell$  have lower bound  $\zeta^2 \frac{\mathcal{M}_1^2 \zeta_1}{N_F}$ .

Localization accuracy is defined as the standard deviation of the molecule's object space localization error. One derivable quantity of interest to researchers will be the amount by which the registration process affects localization of a feature (single molecule) in object space.

**Definition VI.3**—Let  $\Omega_{1,F} = (\mathcal{M}_1^2 \zeta_1 / N_{1,F}) I_2$  be the covariance matrix for the measurement error in localizing a feature at true location  $x_{1,F} \in \mathcal{A}_1$  and let  $\Omega_\ell$  be the covariance matrix of the LRE, then the localization loss ratio,  $\Delta_F$  say, has  $(m, n)$ th element

$$[\Delta_F]_{mn} \equiv \left( \frac{\mathcal{M}_1^2 [\Omega_\ell]_{mn}}{\mathcal{M}_2^2 [\Omega_{1,F}]_{mn}} \right)^{\frac{1}{2}}.$$

From (18) we now give the following result for the registration induced error ratio:

**Proposition VI.4**—Under Assumptions (i)–(iii) the diagonal terms of the registration induced error ratio matrix,  $\Delta_F$ , for a feature at true location  $x_{1,F} \in \mathcal{A}_1$  are given as (for  $n = 1, 2$ )

$$[\Delta_F]_{nn} \approx \left( \frac{\zeta_1^2 \mathcal{M}_1^2}{\mathcal{M}_2^2} + \frac{N_{1,F}}{K} \left( 1 + \left( \frac{r_F}{\kappa} \right)^2 \right) \times \left( \zeta_2^2 \frac{\mathcal{M}_1^2}{\mathcal{M}_2^2} \frac{1}{N_1} + \frac{\zeta_2}{\zeta_1} \frac{1}{N_2} \right) \right)^{\frac{1}{2}}. \quad (19)$$

**Proposition VI.5**—If we wish the localization accuracy of a registered feature in  $\mathcal{A}_2$  to be restricted to  $p\%$  more than the localization accuracy in  $\mathcal{A}_1$ , i.e.  $[\Delta_F] \leq 1 + (p/100)$ , assuming strict equality in (19), we require the following inequality is satisfied:

$$\frac{1}{K} \left( \frac{\mathcal{M}_1^2}{\mathcal{M}_2^2} \frac{1}{N_1} + \frac{\zeta_2}{\zeta_1} \frac{1}{N_2} \right) \leq \frac{1}{N_F \left( 1 + \left( \frac{r_F}{\kappa} \right)^2 \right)} \left( \frac{(1 + \frac{p}{100})^2}{\zeta^2} - 1 \right). \quad (20)$$

Consider a multitemporal registration scenario, i.e. point spread functions, numerical aperture and photon wavelengths can be considered identical  $\zeta_1 = \zeta_2$ ,  $\mathcal{M}_1 = \mathcal{M}_2$  and  $N_{1,k} \approx N_{2,k}$  for all  $k = 1, \dots, K$ . Assume  $A$  is a rotation with no scaling (i.e.  $\zeta = 1$ ) and we have an arbitrary translation  $s$ . Inequality (20) becomes

$$\frac{1}{K} \left( \frac{1}{N_1} + \frac{1}{N_2} \right) \leq \frac{1}{N_{1,F} \left( 1 + \left( \frac{r_F}{\kappa} \right)^2 \right)} \left( \frac{2p}{100} + \left( \frac{p}{100} \right)^2 \right). \quad (21)$$

For example, suppose we image a single molecule in  $\mathcal{A}_1$  with a photon count of  $N_{1,F} = 200$  on the outer corner of a square  $\mathcal{A}_1$  of dimensions  $l \times l$ , then  $r_F^2 = l^2/2$ . Consider CPs that are uniformly distributed in  $\mathcal{A}_1$ , then  $\kappa^2 = l^2/12$  and  $(r_F/\kappa)^2 = 6$ . To restrict the loss in localization accuracy of the single molecule due to the registration process to within 10% then we require  $K^{-1}((1/N_1) + (1/N_2)) \leq 1.5 \times 10^{-4}$ . In such a scenario, 10 CPs with a mean photon count of 1350 would be sufficient. A single molecule at the center of  $\mathcal{A}_1$  would undergo a loss in localization accuracy of only 0.71%.

### C. Non-Diagonal $\Omega_0$

Let us now consider the case where there exists off-diagonal terms in the matrices  $\Omega_{1,0}$  and

$\Omega_{2,0}$ —see Section VI-A.  $\Omega_0$  is still of form (16) with  $\Omega_{j,0} = \begin{bmatrix} \sigma_{j,0}^2 & \rho_{j,0} \\ \rho_{j,0} & \sigma_{j,0}^2 \end{bmatrix}$ . Assuming the analysis in Section VI-B on the order of magnitude of relative terms still holds, and matrix

$$R = \begin{bmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{bmatrix} \quad (21)$$

represents a rotation by angle  $\varphi$ , we have the following

$$[\Omega_\tau]_{11} \approx K^{-1} \left( \zeta^2 \bar{\sigma}_1^2 - \zeta^2 \bar{\rho}_1 \sin(2\varphi) + \bar{\sigma}_2^2 \right) \quad (22)$$

$$[\Omega_\tau]_{22} \approx K^{-1} \left( \zeta^2 \bar{\sigma}_1^2 + \zeta^2 \bar{\rho}_1 \sin(2\varphi) + \bar{\sigma}_2^2 \right) \quad (23)$$

$$[\Omega_\tau]_{12} \approx K^{-1} \left( \zeta^2 \bar{\rho}_1 \cos(2\varphi) + \bar{\sigma}_2^2 \right) \quad (24)$$

where  $\bar{\sigma}_j^2 = \sigma_{j,0}^2 / \bar{N}_j$ ,  $j = 1, 2$ , and  $\bar{\rho}_1 = \rho_{1,0} / \bar{N}_1$ . Equations (22), (23) and (24) show that when the covariance of the CP measurement errors contain off-diagonal terms, the covariance matrices of the TRE and LRE are dependent on the amount of rotation in the transform. For example, for  $\varphi \in (0, \pi/4)$  then  $[\Omega_\tau]_{11} < [\Omega_\tau]_{22}$ , and for  $\varphi \in (-\pi/4, 0)$  we have  $[\Omega_\tau]_{11} > [\Omega_\tau]_{22}$ . However, for small rotation angles  $\varphi$  the effect of off diagonal terms is negligible and the results in Propositions VI.2 and VI.4 will still be appropriate.

#### D. Geometrically Regular Control Point Configurations

In addition to control points that appear to the experimenter at random locations in the image space, it is also useful to consider the use of deterministic, geometrically regular CP locations set by the experimenter. These could take the form of a grid or ellipse (which includes the special case of a circle, provided pair correspondence between the CPs in both images can be guaranteed). We once again take the weighting factor for the covariance matrix (the reciprocal of the photon count) associated with a CP to be a realization of a random variable  $\mathcal{N}$  and is completely independent of CP location.

Consider a square grid of side length  $a$  centered at the origin, where CPs are evenly spaced including positioning at each vertex (i.e.  $K$  is always square). It can be shown  $\Psi = E\{\mathcal{N}^{-1}\} (a^2/3)I_2$ . In this situation (18) is valid, replacing  $\kappa$  with  $a^2/3$ .

We also consider an ellipse centered at the origin with the major axis running along the  $x$  axis with major radius  $a$  and minor radius  $b$ , then it can be shown that  $\Psi = (1/2) E\{\mathcal{N}^{-1}\} \text{diag}\{a^2, b^2\}$ . Equation (18) is still valid, however for the case  $n = m = 1$  then  $\kappa$  should be replaced by  $a$ . When  $n = m = 2$ ,  $k$  is replaced  $b$ . With  $a > b$ , this implies that image registration is better resolved in the direction of the major axis as opposed to the direction of the minor axis.

## VII. Simulations

Here we seek verification of the results in Section VI through Monte Carlo simulations. We consider a multitemporal registration scenario.  $\mathcal{A}_1$  and  $\mathcal{A}_2$  each comprise of  $512 \times 512$  pixels, with each pixel being of dimensions  $16 \mu\text{m} \times 16 \mu\text{m}$ . The system magnification is  $\mathcal{M}_1 = \mathcal{M}_2 = 100$ , so in object space each pixel corresponds to a square of  $0.16 \mu\text{m} \times 0.16 \mu\text{m}$ . We consider the affine transformation  $T: \mathcal{A}_1 \rightarrow \mathcal{A}_2$ ,  $T(x) = Ax + s$  with  $A$  being a rotation of form (21) with  $\varphi = \pi/6$  and  $s = [480, 480]^T$  (corresponding to 30 pixels in each direction).

We assume the measurement errors for the bead positions in  $\mathcal{A}_j$  have mean zero and covariance matrix  $(\mathcal{M}_j \zeta_j / N_{j,k}) I_2$  where  $\zeta_j = \lambda_{j,em}^2 / (4\pi^2 n_F^2)$  [26].  $\lambda_{j,em}$  is the wavelength of the photons observed in the  $j$ th image and is set as  $0.520 \mu\text{m}$  for  $j = 1$  and  $2$ .  $n_F$  is the numerical aperture of the optical system and set to a typical value of  $1.4$  for both images.  $N_{j,k}$  is the photon count associated with the  $k$ th CP in the  $j$ th image. We consider a single molecule in  $\mathcal{A}_1$  with true position  $x_{1,F} = (1600, 1200)$  (giving  $r_F = \|x_{1,F}\| = 2000 \mu\text{m}$ ), assuming localization is subject to measurement error with covariance  $\sigma_F^2 I_2$ , where  $\sigma_F = 0.186 \mu\text{m}$ . Estimators  $\hat{A}$  and  $\hat{s}$  were computed as outlined in Theorem III.3.



Simulations for a multimodal scenario would be carried out in an analogous manner. However, instead of the system parameters (photon wavelength, numerical aperture, system magnification) being the same for both images, they would in general differ between  $\mathcal{A}_1$  and  $\mathcal{A}_2$ . These different values can be easily accounted for by calculating the covariance matrix  $\Omega_0$  appropriately and generating measurement errors with this covariance. LRE and TRE covariance terms can be calculated by substituting the system parameters into (17) and (18). Affine transformation parameters  $A$  and  $s$  are then estimated in an identical way.

### A. Distributional Analysis

Fig. 2 is a quantile-quantile (Q-Q) plot for the distribution of the first element of the TRE. The curve is produced by ordering 10000 independent and normalized estimates (with respect to the theoretical variance) into increasing order of the size. The probability of a value less than the  $j$ th ordered estimate (or sample quantile) is  $p_j = j/10001$  to a close approximation. The corresponding theoretical quantile of the normal distribution is the value  $v_j$  such that  $p_j = F(v_j)$ , where  $F(\cdot)$  is the cumulative distribution function of the normal distribution. The values  $v_1, \dots, v_{10000}$  are plotted on the  $x$ -axis against the ordered estimates for (a)  $K = 6$ , (b) 10, (c) 15, (d) 20 uniformly distributed CPs. Even for low numbers of CPs the fit to the derived “large  $K$ ” distribution is striking, and similar results were seen for the second element of the TRE.

We note that the LRE  $\ell_F$  is comprised of contributions from the TRE  $\tau$  and the measurement error  $\epsilon_{1,F}$ . The elements of the TRE have standard deviation of order  $\frac{2}{\sqrt{K}}$  nm, whereas the elements of  $\epsilon_{1,F}$  have standard deviation of order 2 nm. The LRE ( $\sim 2 + \left(\frac{1}{\sqrt{K}}\right)$  nm) is therefore dominated by the Gaussian measurement error  $\epsilon_{1,F}$ . For this reason it is more meaningful to focus analysis on the distribution of the TRE.

### B. Localization Analysis

We now consider four different CP configurations; CP locations are (a) normally distributed about the center of  $\mathcal{A}_1$  with covariance  $\kappa^2 I_2$ ,  $\kappa = 1350 \mu\text{m}$ , (b) uniformly distributed in  $\mathcal{A}_1$ , (c) arranged in an ellipse with major radius  $4000 \mu\text{m}$  and minor radius  $3000 \mu\text{m}$ , and (d) arranged in a square grid of side  $8100 \mu\text{m}$ . Due to restrictions on the possible values of  $K$  in (d) we consider simulations where  $K$  takes the value of the square numbers from 4 to 64 inclusive. For configurations (a)–(d) the sample standard deviation of  $[\ell_F]_1$  for 100000 Monte Carlo simulations is computed for each value of  $K$ , while keeping the same uniform distribution that the CP photon counts are sampled from. The results are plotted in Fig. 3, as are the values predicted by (18). The dashed line marks the theoretical lower bound

$$\sigma_F = \left( \left[ \Omega_{1,F} \right]_{11} \right)^{\frac{1}{2}} \text{ of the LRE standard deviation.}$$

The accuracy for normally and uniformly distributed CPs (scenarios (a) and (b)) shows some deviation from the large sample results for small  $K \sim 9$ , and in such a circumstance (20) is not totally appropriate. A better interpretation is; at the very least this inequality must be satisfied. However, for larger values of  $K$  this provides an excellent guide to experimenters. The elliptical and grid configurations (scenarios (c) and (d)) show a much closer fit for small  $K$  compared to the random configurations (scenarios (a) and (b)). This is most likely due to the random positioning of the CPs adding to the overall variance of the parameter estimators. In the deterministic case (20) is an excellent guide to experimenters.

In Fig. 4 we keep the number of CPs constant at  $K = 16$  and change the mean of the uniform distribution from which the photon counts are sampled. For configurations (a)–(d), the

sample standard deviation of the first element of the LRE for  $\mathcal{L}_F$  100000 Monte Carlo simulations is computed. We also plot the values as predicted by (18). The dashed line marks the theoretical lower bound  $\sigma_F = ([\Omega_{1,F}]_{11})^{\frac{1}{2}}$  of the LRE standard deviation. By increasing the photon count, the asymptotic results can be readily achieved and the key results presented in this paper provide an excellent guide.

We have shown in Section VI-B that for a diagonal covariance matrix  $\Omega_0$ , as is the case in our simulations, then the LRE and TRE are independent of rotation angle and translation in the affine transformation. In Fig. 5 we keep the number of CPs constant at  $K = 16$ , the mean of the uniform distribution from which the photon counts are sampled is kept constant and with affine transformation parameter  $A = R$  of form (21) change the angle of rotation  $\phi$ . For each rotation angle the translation is by a random amount. For configurations (a)–(d), the sample standard deviation of the first element of the LRE  $\mathcal{L}_F$  for 100000 Monte Carlo simulations is computed. We also plot the values as predicted by (18). The dashed line marks the theoretical lower bound  $\sigma_F = ([\Omega_{1,F}]_{11})^{\frac{1}{2}}$  of the LRE standard deviation. It is clear that the standard deviation of the LRE is invariant to rotation angle and translation, as predicted.

**C. Non-Gaussian Measurement Errors**—The derived Gaussian distribution for the TRE presented in this paper is given under the assumption that measurement errors in localizing CPs are themselves Gaussian distributed. The measurement errors in a microscopy experiment are often assumed to be Gaussian—which is justified by large sample results and evidence based on the analysis of experimental and simulated data—it is nevertheless useful to know how robust the results of this paper are under deviations away from this Gaussian assumption. We therefore consider the TRE when measurement errors are uniformly distributed, an extreme deviation from Gaussianity.

Fig. 6 gives Q-Q plots to compare the first element of the TRE against the normal distribution. Labels (a)–(d) refer to the same values of  $K$  as with Fig. 2. Even with such a pronounced change in the distribution of the CP measurement errors the plots show the normal distribution is still an appropriate approximation for the TRE, although the exact distribution is unknown.

## VIII. Imaging Data and Method Comparison

We now apply the algorithm presented in this paper to real imaging data. Analysis has been conducted by performing image registration between a pair of images of the same object space taken by two separate cameras (multimodal registration). We have 599 repeat captures, resulting in 599 pairs of images to register. The registration was performed using 27 TetraSpeck fluorescent beads visible in both fields of view. System parameters were identical for each image, with system magnifications of 63, numerical apertures equal to 1.45 and the photon wavelength distributions peaking at 638 nm.

To calculate the TRE for each of the 599 registrations we need to know the exact coordinates of a point in each image that perfectly map to one another under the affine transformation. Without knowledge of the true transformation parameters this is not available and we instead isolate one of the beads and average its location over the 599 images for each camera. This will give a very high precision estimate of its true location in each image and these coordinates are used as the reference points for the TRE. The TRE is then calculated for each of the 599 registration experiments using two different methods.

The first method is the weighted covariance GLS method presented in this paper (developed in Section III-A and implemented on simulations in Section VII). This method uses the system parameters quoted above and the point spread function to calculate covariance matrix  $\Omega_0$  and takes into account both the photon count and location estimates for each bead to compute estimates of the affine transformation parameters  $A$  and  $s$ —see Fig. 1 for details.

The second method used is the traditional GLS method presented in [22, page 28] (that assumes homoscedastic measurement errors and is equivalent to the ML approach to image registration used by [23]). This method uses just the location estimates for each bead in estimating  $A$  and  $s$ .

For both methods, the sample standard deviation of the TRE for the  $x$ - and  $y$ -direction is computed and displayed in Table I. This was repeated 27 times, each time isolating a different bead to act as the reference point.

The results show the weighted covariance GLS methods presented here consistently outperforms the traditional GLS method. Comparisons with the theoretical results would be misleading because of a lack of access to the ground truth and photon counts significantly differ across the 599 samples.

## IX. Concluding Remarks

We have used a heteroscedastic generalized least squares approach to point-based image registration. This has allowed an asymptotic analysis of the distributional properties of the TRE and the LRE; a new measure of registration success that is of interest in microscopy. By considering the weighted covariance we have derived closed form expressions for both the registration estimators and the large sample TRE/LRE distributions. These distributions can be used for determining confidence intervals of errors induced by registration.

Fluorescence microscopy image registration was used as a motivating example. Here we have derived the TRE and LRE covariance matrices in terms of the number of CPs, their spread in the object space, and associated photon counts. The relative loss in localization accuracy of a imaged single molecule was further derived. When the covariance matrices of the CP measurement errors are multiples of the identity then there is no dependence on the translation or rotation components of the affine transformation. When they have off-diagonal terms then the TRE/LRE are dependent on the rotation part of the affine transformation but still remain independent of the translation. Theoretical results have been verified with Monte Carlo simulations and even for relatively small values of  $K$  simulations show excellent agreement with the theory. However, small but notice able discrepancies between the simulations and large sample expressions do occur for small  $K$  and in such a situation the results presented here should be considered a useful guide rather than exact.

Using real imaging data the weighted covariance GLS method presented here has been shown to consistently outperform the traditional GLS method.

## Acknowledgments

The authors would like to thank Dr. S. Ram, D. Kim and Prof. E. Sally Ward for the data analyzed in Section VIII.

The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Yufei Huang. This work was supported in part by the National Institute of Health grant R01 GM085575.

## Appendix A

Table II gives a list of standard abbreviations. Table III gives a list of common operators and functions. Table IV gives a list of standard notations. Table V gives a list of fluorescence microscopy notations.

## Appendix B

### Proof of Theorem III.2

Under Assumption II, the 4D multivariate Gaussian pdf for measurement vector

$y_k = [y_{1,k}^T, y_{2,k}^T]^T$  is given as

$$p(y_k) = (2\pi)^{-\frac{1}{2}} |\Omega_k|^{-\frac{1}{2}} e^{-\frac{1}{2} (y_k - \mu_k)^T \Omega_k^{-1} (y_k - \mu_k)}$$

where  $\mu_k = \Lambda x_{1,k} + \alpha$ ,  $k = 1, \dots, K$ ,  $\Lambda = [I_d, A^T]^T$ . The likelihood function  $\mathcal{L}(A, s, X_1; Y)$  for  $A$ ,  $s$  and  $X_1$  is given as

$$(2\pi)^{-K/2} \left[ \prod_{k=1}^K |\Omega_k|^{-\frac{1}{2}} \right] e^{-\frac{1}{2} \sum_{k=1}^K (y_k - \mu_k)^T \Omega_k^{-1} (y_k - \mu_k)}.$$

The parameter values that maximize  $\mathcal{L}(A, s, X_1; Y)$  are those that minimize the term

$\sum_{k=1}^K (y_k - \mu_k)^T \Omega_k^{-1} (y_k - \mu_k)$ . Noticing

$$\sum_{k=1}^K (y_k - \mu_k)^T \Omega_k^{-1} (y_k - \mu_k) = \sum_{k=1}^K r_k^T \Omega_k^{-1} r_k = r,$$

it follows that  $\hat{A}_{ml} = \hat{A}_{ols}$ ,  $\hat{s}_{ml} = \hat{s}_{ols}$  and  $\hat{X}_1^{ml} = \hat{X}_1^{ols}$ .

### Lemma B.1

Let the  $k$ th column of  $X_1$  be given as

$$\hat{x}_{1,k} = \left( \Lambda^T \Omega_k^{-1} \Lambda \right)^{-1} \Lambda^T \Omega_k^{-1} (y_k - \alpha), \quad (25)$$

then for any fixed  $A$  and  $s$ ,  $r(s, A, X_1; Y) = r(s, A, X_1; Y)$  and  $\mathcal{L}(A, s, X_1; Y) = \mathcal{L}(A, s, X_1; Y)$ .

**Proof**—Let  $r_k \equiv (y_k - \mu_k)^T \Omega_k^{-1} (y_k - \mu_k)$ , then

$$\frac{\partial r}{\partial x_{1,k}} = \frac{\partial r_k}{\partial x_{1,k}} = -2\Lambda \Omega_k^{-1} y_k + 2\Lambda^T \Omega_k^{-1} \Lambda x_{1,k} + 2\Lambda^T \Omega_k^{-1} \alpha. \text{ The conditional estimator } x_{1,k}^{\hat{}} \text{ is}$$

the value of  $x_{1,k}$  that gives  $\frac{\partial r}{\partial x_{1,k}} = \frac{\partial r_k}{\partial x_{1,k}} = 0$ . This gives  $\hat{x}_{1,k} = \left( \Lambda^T \Omega_k^{-1} \Lambda \right)^{-1} \Lambda^T \Omega_k^{-1} (y_k - \alpha)$  and the result follows.

The ML estimators (under Assumption II) and OLS estimators (under Assumption I) for  $A$  and  $s$  are the values that minimize  $r(s, A, \hat{X}_1; Y) = \sum_{k=1}^K \tilde{r}_k^T \tilde{r}_k$  where  $\tilde{r}_k = \Omega_k^{-\frac{1}{2}} r_k$ . Substituting in (25) gives

$$\tilde{r}_k = \left( I - \Omega_k^{-\frac{1}{2}} \Lambda (\Lambda^T \Omega_k^{-1} \Lambda)^{-1} \Lambda^T \Omega_k^{-\frac{1}{2}} \right) \Omega_k^{-\frac{1}{2}} (y_k - \alpha). \quad (26)$$

We note that  $\left[ \Omega_k^{-\frac{1}{2}} \Lambda, \Omega_k^{\frac{1}{2}} U^T \right]$  is a non-singular square matrix and  $\left( \Omega_k^{-\frac{1}{2}} \Lambda \right)^T \Omega_k^{\frac{1}{2}} U^T = \Lambda^T U^T = 0$ .

### Lemma B.2

Let  $F$  and  $G$  be equal dimension  $n \times p$  matrices such that the matrix  $[F, G]$  is square nonsingular and  $F^T G = 0$ , then  $I_n = F(F^T F)^{-1} F^T + G(G^T G)^{-1} G^T$ .

Lemma (B.2) gives

$$I - \Omega_k^{-\frac{1}{2}} \Lambda (\Lambda^T \Omega_k^{-1} \Lambda)^{-1} \Lambda^T \Omega_k^{-\frac{1}{2}} = \Omega_k^{\frac{1}{2}} U^T \Phi_k^{-1} U \Omega_k^{\frac{1}{2}}. \quad (27)$$

Substituting (27) into (26) gives  $\tilde{r}_k = \Gamma_k(A)^T q_k$  where  $\tilde{q}_k = \Phi_k^{-\frac{1}{2}} q_k$  and  $\Gamma_k(A) \equiv \Phi_k^{-\frac{1}{2}} U \Omega_k^{\frac{1}{2}}$  is a row orthogonal matrix. Hence  $\Gamma_k(A) \Gamma_k(A)^T = I_{2d}$  and  $\tilde{r}_k^T \tilde{r}_k = \tilde{q}_k^T \tilde{q}_k$ , giving

$$r(s, A, \hat{X}_1; Y) = q(s, A; Y)$$

under Assumption I.

## Appendix C

### Proof of Theorem III.3

This proof borrows from [22]. Consider the Frobenius norm

$$\|Q(s, A; Y)\|_F^2 = \sum_{k=1}^K q_k^T \Phi_k^{-1} q_k.$$

Then

$$\frac{\partial}{\partial s} \|Q(s, A; Y)\|_F^2 = 2 \sum_{k=1}^K \Phi_k^{-1} U y_k - 2 \sum_{k=1}^K \Phi_k^{-1} s$$

and consequently by setting equal to zero we conclude for any fixed  $A$   $\|Q(s, A; Y)\|_F^2 \geq \|Q(\hat{s}, A; Y)\|_F^2$  where

$$\hat{s} = \left( \sum_{k=1}^K \Phi_k^{-1} \right)^{-1} \sum_{k=1}^K \Phi_k^{-1} U y_k. \quad (28)$$

Under Assumption IV we have  $\Phi_k = \eta_k \Phi_0$  and we can simplify (28) to give  $\hat{s} =$

$$\gamma^{(K)-1} U \bar{y}^{(K)}, \text{ where } \gamma^{(K)} \equiv \left( \frac{1}{K} \right) \sum_{k=1}^K \eta_k^{-1} \text{ and } \bar{y}^{(K)} \equiv \left( \frac{1}{K} \right) \sum_{k=1}^K \eta_k^{-1} y_k.$$

Define  $v_k \equiv y_k - \gamma^{(K)-1} \bar{y}$ ,  $k = 1, \dots, K$  and the matrix  $V \equiv [v_1, \dots, v_K]$ , then it can be shown for any fixed  $A$  that

$$\|Q(s, A; Y)\|_F^2 \geq \|Q(\hat{s}, A; Y)\|_F^2 = \|Q(0, A; V)\|_F^2 \quad (29)$$

where

$$Q(0, A; V) = \left[ \Phi_k^{-\frac{1}{2}} U v_1, \dots, \Phi_k^{-\frac{1}{2}} U v_K \right]$$

with equality in (29) holding if and only if

$$s = \hat{s}(A) = \gamma^{(K)-1} U \bar{y}.$$

We now note that under Assumption IV

$$Q(0, A; V) = \left[ \Gamma(A) \Omega_1^{-\frac{1}{2}} v_1, \dots, \Gamma(A) \Omega_K^{-\frac{1}{2}} v_K \right],$$

where

$$\Gamma(A) \equiv \Phi_k^{-\frac{1}{2}} U \Omega_k^{\frac{1}{2}} = \Phi_0^{-\frac{1}{2}} U \Omega_0^{\frac{1}{2}},$$

and hence we can write  $Q(0, A; V) = \Gamma(A) \Omega_0^{-\frac{1}{2}} \tilde{V}$  with  $\tilde{V} = \left[ \eta_1^{-\frac{1}{2}} v_1, \dots, \eta_K^{-\frac{1}{2}} v_K \right]$ . We note  $\tilde{V} \tilde{V}^T = W^{(K)}$ .

We are required to minimize

$$\|Q(0, A; V)\|_F^2 = \|\Gamma(A) \Omega_0^{\frac{1}{2}} \tilde{V}\|_F^2$$

with respect to  $A$ .

### Lemma C.1

Let  $M \in \mathbb{R}^{m \times n}$  with  $m \geq n$  and  $\text{rank}(M) = r$  and let  $P \in \mathbb{R}^{n \times q}$ ,  $q \geq r$ , be a matrix with  $q$  orthonormal columns. Then  $\|MP\|_F^2 = \text{tr}(P^T M^T M P)$  is minimized for given  $M$ , when  $P = G_k$  where  $U S G^T$  is a singular value decomposition (svd) of  $M$  and  $G_k = [g_{m-k+1}, \dots, g_m]$  denotes the matrix  $k$  of orthonormal right singular vectors of  $M$  corresponding to the  $k$  smallest singular values.

We note that

$$\|\Gamma(A)\Omega_0^{-\frac{1}{2}}\tilde{V}\|_F^2 = \|\tilde{V}^T\Omega_0^{-\frac{1}{2}}\Gamma^T(A)\|_F^2,$$

and  $\Gamma^T(A)$  is a  $2d \times d$  matrix with orthonormal columns. Let  $U S G^T$  be the svd of  $\tilde{V}^T\Omega_0^{-\frac{1}{2}}$  where we partition the matrix of right singular vectors

$$\tilde{G} = \begin{bmatrix} \tilde{G}_{11} & \tilde{G}_{12} \\ \tilde{G}_{21} & \tilde{G}_{22} \end{bmatrix}.$$

From Lemma C.1 we have

$$\|Q(0, A; V)\|_F^2 = \text{tr}\left(\Gamma(A)\Omega_0^{-\frac{1}{2}}\tilde{V}\tilde{V}^T\Omega_0^{-\frac{1}{2}}\Gamma^T(A)\right) \geq \text{tr}\left(\Omega_0^{-\frac{1}{2}}\tilde{V}\tilde{V}^T\Omega_0^{-\frac{1}{2}}\right),$$

with equality reached when  $\hat{A}$  is such that  $\Gamma(\hat{A}) = [\tilde{G}_{12}^T, \tilde{G}_{22}^T]$ . From the svd of  $\tilde{V}^T\Omega_0^{-\frac{1}{2}}$  we have

$$W^{(K)}\Omega_0^{-1} = \tilde{V}\tilde{V}^T\Omega_0^{-1} = \Omega_0^{\frac{1}{2}}\tilde{G}S^T S\tilde{G}^T\Omega_0^{-\frac{1}{2}} = GEG^{-1},$$

the eigendecomposition of  $W^{(K)}\Omega_0^{-1}$  where the columns of  $G = \Omega_0^{\frac{1}{2}}\tilde{G}$  are the eigenvectors of  $W^{(K)}\Omega_0^{-1}$ , with inverse

$$G^{-1} = \tilde{G}^T\Omega_0^{-\frac{1}{2}}. \quad (30)$$

We have  $\Gamma(\hat{A}) = [\tilde{G}_{12}^T, \tilde{G}_{22}^T]$  which gives  $\Phi_0^{-\frac{1}{2}}U\Omega_0^{\frac{1}{2}} = [\tilde{G}_{12}^T, \tilde{G}_{22}^T]$ , resulting in

$$\Phi_0^{-\frac{1}{2}}U = [\tilde{G}_{12}^T, \tilde{G}_{22}^T]\Omega_0^{-\frac{1}{2}} = [[G^{-1}]_{21}, [G^{-1}]_{22}],$$

with the final equality coming from (30), where we make the partition

$$G^{-1} = \begin{bmatrix} [G^{-1}]_{11} & [G^{-1}]_{12} \\ [G^{-1}]_{21} & [G^{-1}]_{22} \end{bmatrix}.$$

With

$$\Phi_0^{-\frac{1}{2}} U = \left[ -\Phi_0^{-\frac{1}{2}} \hat{A}, \Phi_0^{-\frac{1}{2}} \right] = \left[ [G^{-1}]_{21}, [G^{-1}]_{22} \right],$$

we have  $\Phi_0^{-\frac{1}{2}} = [G^{-1}]_{22}$  and  $\hat{A} = -([G^{-1}]_{22})^{-1} [G^{-1}]_{21}$ . Using the block inverse of  $G$  we have  $[G^{-1}]_{21} = -\left(G_{22} - G_{21} G_{11}^{-1} G_{12}\right)^{-1} G_{21} G_{11}^{-1}$  and  $[G^{-1}]_{22} = \left(G_{22} - G_{21} G_{11}^{-1} G_{12}\right)^{-1}$ , and it follows that  $\hat{A} = -\left([G^{-1}]_{22}\right)^{-1} [G^{-1}]_{21} = G_{21} G_{11}^{-1}$ . We have now shown

$$\|Q(s, A; Y)\|_F \geq \|Q(\hat{s}, \hat{A}; Y)\|_F$$

for all  $A$  and  $s$ .

We are left with showing consistency. Let

$$W^{(K)} \equiv \left(\frac{1}{K}\right) \sum_{k=1}^K \eta_k^{-1} \left(y_k - \gamma^{(K)-1} \bar{y}^{(K)}\right) \left(y_k - \gamma^{(K)-1} \bar{y}^{(K)}\right)^T$$

and consider eigendecomposition  $W^{(K)} \Omega_0^{-1} \equiv G^{(K)} E^{(K)} G^{(K)-1}$  with  $E^{(K)} = \text{diag}\{e_1, e_2, \dots, e_{2d}\}$  where  $e_1 \dots e_{2d}$  are the ordered eigenvalues of  $W^{(K)} \Omega_0^{-1}$  and the columns of  $G^{(K)}$  are the corresponding eigenvectors. Under Assumptions I, IV and V(i) it can be shown

$W \Omega_0^{-1} \equiv \lim_{K \rightarrow \infty} W^{(K)} \Omega_0^{-1}$  exists and is equal to  $W \Omega_0^{-1} = \Lambda \Psi \Lambda^T \Omega_0^{-1} + I_{2d}$  wp1, where

$\Psi = \lim_{K \rightarrow \infty} \Psi^{(K)}$ ,  $\Psi^{(K)} \equiv \Xi^{(K)} - \gamma^{(K)-1} \bar{x}_1^{(K)} \bar{x}_1^{(K)T}$ ,  $\Xi^{(K)} \equiv \left(\frac{1}{K}\right) \sum_{k=1}^K \eta_k^{-1} x_{1,k} x_{1,k}^T$  and  $\Lambda = [I_d, A^T]^T$ . Further to this, Assumption V(ii) implies  $W \Omega_0^{-1}$  is positive definite. Under these conditions we give the following as a summary to the results found in [22, Section 3].

### Lemma C.2

$W \Omega_0^{-1}$  will have eigendecomposition  $W \Omega_0^{-1} = \mathcal{G} E \mathcal{G}^{-1}$  where  $\mathcal{G} = \lim_{K \rightarrow \infty} G^{(K)}$  and  $E = \lim_{K \rightarrow \infty} E^{(K)}$  each wp1.

Let  $\gamma_1 \ \gamma_2 \ \dots \ \gamma_d = 0$  be the eigenvalues of  $\Psi \Lambda^T \Omega_0^{-1} \Lambda$ .  $\Psi \Lambda^T \Omega_0^{-1} \Lambda$  is symmetric hence there exists  $d \times d$  matrix  $\psi$  such that  $\Psi \Lambda^T \Omega_0^{-1} \Lambda = \psi D_\gamma \psi^{-1}$  where  $D_\gamma = \text{diag}\{\gamma_1, \gamma_2, \dots, \gamma_d\}$ . We therefore have



$$\mathcal{G}_1 \Lambda^T \Omega_0^{-1} \Lambda \psi = \psi D_\gamma$$

and hence

$$W \Omega_0^{-1} \Lambda \psi = \left( I_{2d} + \Lambda \Psi \Lambda^T \Omega_0^{-1} \right) \Lambda \psi = \Lambda \psi (I_d + D_\gamma).$$

Consequently, we conclude that  $\Lambda \psi$  are ordered eigenvectors of  $W \Omega_0^{-1}$  with eigenvalues  $E = (I_d + D_\gamma)$ . By also considering  $W \Omega_0^{-1} \Gamma^T(A) = \Gamma^T(A)$  we see that the columns of  $\Lambda \psi$  are the  $d$  eigenvectors with the largest corresponding eigenvalues, and the columns of  $\Gamma^T(A)$  are the eigenvectors with the smallest eigenvalues (each equal to one). As such, if we define  $\mathcal{G}$  to be a matrix whose columns are the eigenvectors of  $W \Omega_0^{-1}$ , ordered by descending size of their corresponding eigenvalues, then

$$\left[ \mathcal{G}_{11}^T, \mathcal{G}_{21}^T \right]^T = \left[ I_d, A^T \right]^T \psi \mathcal{U}$$

for some unitary matrix  $\mathcal{U}$ . From Lemma C.2 this gives

$$\lim_{K \rightarrow \infty} \hat{A} = \lim_{K \rightarrow \infty} G_{21}^{(K)} G_{11}^{(K)-1} = \mathcal{G}_{21} \mathcal{G}_{11}^{-1} = A$$

wp1.

Using this result and under Assumption V  $\lim_{K \rightarrow \infty} \hat{s} = \gamma^{-1} U \bar{y}$  where  $\gamma = \lim_{K \rightarrow \infty} \gamma^{(K)}$  and  $\bar{y} = \lim_{K \rightarrow \infty} \bar{y}^{(K)}$ . By the strong law of large numbers (SLLN),  $\lim_{K \rightarrow \infty} \left( \frac{1}{K} \right) \sum_{k=1}^K \eta_k^{-1} \epsilon_k = 0$  wp1, and hence  $\bar{y} = x \bar{w}$  where  $\bar{x} = \lim_{K \rightarrow \infty} \bar{x}^{(K)}$  with  $\bar{x}^{(K)} = \left( \frac{1}{K} \right) \sum_{k=1}^K \eta_k^{-1} x_k$ . Therefore, with  $U x_k = x_{2,k} - A x_{1,k} = s$  for all  $k = 1, \dots, K$ ,  $\gamma^{-1} U \bar{y} = \gamma^{-1} U x = \gamma^{-1} \gamma s$  and  $\lim_{K \rightarrow \infty} \hat{s} = s$  wp1.

## Appendix D

### Proof of Corollary IV.2

Under Assumptions II, IV, V and IV, and from Theorems III.3 and IV.1, we have that asymptotically in  $K$ ,  $\text{cov}\{K^{\frac{1}{2}} \Delta \beta\} = H^{-1} P (H^{-1})^T$  where  $H$  and  $P$  are given by (6) and (7), respectively. Let us define scalar  $\gamma^{(K)} \equiv \sum_{k=1}^K \eta_k^{-1}$ , vector  $\bar{x}_1^{(K)} \equiv \left( \frac{1}{K} \right) \sum_{k=1}^K \eta_k^{-1} x_{1,k}$  and matrix  $\Xi^{(K)} \equiv \left( \frac{1}{K} \right) \sum_{k=1}^K \eta_k^{-1} x_{1,k} x_{1,k}^T$ . Under Assumption IV it can be shown

$$\begin{aligned}
& \left(\frac{1}{K}\right) \sum_{k=1}^K \Phi_k^{-1} = \gamma^{(K)} \Phi_0^{-1}, \\
& - \left(\frac{1}{K}\right) \sum_{k=1}^K \left(x_{1,k} \otimes \Phi_k^{-1}\right) = -\bar{x}_1^{(K)} \otimes \Phi_0^{-1}, \\
& - \left(\frac{1}{K}\right) \sum_{k=1}^K \left(x_{1,k}^T \otimes \Phi_k^{-1}\right) = -\bar{x}_1^{(K)T} \otimes \Phi_0^{-1} \\
& - \left(\frac{1}{K}\right) \sum_{k=1}^K \left(x_{1,k} x_{1,k}^T \otimes \Phi_k^{-1}\right) = -\Xi^{(K)} \otimes \Phi_0^{-1}
\end{aligned}$$

and

$$\left(\frac{1}{K}\right) \sum_{k=1}^K \left( \left( \Lambda^T \Omega_k^{-1} \Lambda \right)^{-1} + x_{1,k} x_{1,k}^T \right) \otimes \Phi_k^{-1} = \left( \left( \Lambda^T \Omega_0^{-1} \Lambda \right)^{-1} + \Xi^{(K)} \right) \otimes \Phi_0^{-1}.$$

Considering the forms of  $H_0^{(K)}$  and  $P_0^{(K)}$  in (8) and (9), respectively, we write  $H^{(K)} = H_0^{(K)} \otimes \Phi_0^{-1}$  and  $P^{(K)} = P_0^{(K)} \otimes \Phi_0^{-1}$ . Using the mixed-product property of the Kronecker product and the symmetry of  $H$ , then

$$H^{(K)-1} = \left( H^{(K)-1} \right)^T = H_0^{(K)-1} \otimes \Phi_0$$

and

$$B^{(K)} = H_0^{(K)-1} P_0^{(K)} H_0^{(K)-1} \otimes \Phi_0.$$

Using the block matrix inverse

$$\begin{aligned}
\left[ H_0^{(K)-1} \right]_{11} &= \gamma^{(K)-1} + \gamma^{(K)-2} \bar{x}_1^{(K)T} \Psi^{(K)-1} \bar{x}_1^{(K)} \\
\left[ H_0^{(K)-1} \right]_{12} &= \gamma^{(K)-1} \bar{x}_1^{(K)T} \Psi^{(K)-1} \\
\left[ H_0^{(K)-1} \right]_{21} &= \gamma^{(K)-1} \Psi^{(K)-1} \bar{x}_1^{(K)} \\
\left[ H_0^{(K)-1} \right]_{22} &= \Psi^{(K)-1}
\end{aligned}$$

where  $\Psi^{(K)} = \Xi^{(K)} - \gamma^{(K)-1} \bar{x}_1^{(K)} \bar{x}_1^{(K)T}$ . Results follow by computing  $H_0^{(K)-1} P_0^{(K)} H_0^{(K)-1}$ , forming the Kronecker product with  $\Phi_0$  and taking limits as  $K \rightarrow \infty$  under Assumption V.

## Appendix E

### Proof of Lemma VI.1

$\bar{x}_1 \equiv \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{k=1}^K \eta_k^{-1} x_{1,k}$ . With weighting factors  $\{\eta_k, k = 1, \dots, K\}$  being independent of points  $\{x_{1,k}, k = 1, \dots, K\}$  by SLLN

$$\bar{x}_1 = E \left\{ \mathcal{N}^{-1} \mathcal{X} \right\} = E \left\{ \mathcal{N}^{-1} \right\} E \left\{ \mathcal{X} \right\} = E \left\{ \mathcal{N}^{-1} \right\} \mu$$

wp1. Let us consider the matrix

$$\Psi = \Xi - \gamma^{-1} \bar{x}_1 \bar{x}_1^T,$$

where  $\gamma = \lim_{K \rightarrow \infty} \left( \frac{1}{K} \right) \sum_{k=1}^K \eta_k^{-1}$  and  $\Xi \equiv \sum_{k=1}^K \eta_k^{-1} x_{1,k} x_{1,k}^T$ . By SLLN  $\gamma = E \left\{ \mathcal{N}^{-1} \right\}$ . By SLLN and independence

$$\Xi = E \left\{ \mathcal{N}^{-1} \mathcal{X} \mathcal{X}^T \right\} = E \left\{ \mathcal{N}^{-1} \right\} E \left\{ \mathcal{X} \mathcal{X}^T \right\}.$$

Hence

$$\Psi = E \left\{ \mathcal{N}^{-1} \right\} \left( E \left\{ \mathcal{X} \mathcal{X}^T \right\} - E \left\{ \mathcal{X} \right\} E \left\{ \mathcal{X} \right\}^T \right) = E \left\{ \mathcal{N}^{-1} \right\} \Pi$$

wp1.

## References

1. Zitová B, Flusser J. Image registration methods: A survey. *Image Vis Comput.* 2003; 21:977–1000.
2. Chumchob N, Chen K. A robust affine image registration method. *Int J Numer Anal Mod.* 2009; 6:311–334.
3. Myronenko A, Song X. Intensity-based image registration by minimizing residual complexity. *IEEE Trans Med Imag.* 2010; 29, no. 11:1882–1891.
4. Liao S, Chung ACS. Feature based nonrigid brain MR image registration with symmetric alpha stable filters. *IEEE Trans Med Imag.* 2010; 29, no. 1:106–119.
5. Yasein MS, Agathoklis P. A feature-based image registration technique for images of different scales. *IEEE Int Symp Circ.* 2008:3558–3560.
6. Li W, Leung H. A maximum likelihood approach for image registration using control point and intensity. *IEEE Trans Image Process.* 2004; 13, no. 8:1115–1127. [PubMed: 15326853]
7. Bates M, Huang B, Dempsey GT, Zhuang X. Multicolor super-resolution imaging with photo-switchable fluorescent probes. *Sci.* 2007; 317:1749–1753.
8. Stirling Churchman L, Ökten RS, Rock J, Dawson JF, Spudich JA. Single molecule high-resolution colocalization of Cy3 and Cy5 attached to macromolecules measures intramolecular distances through time. *PNAS.* 2005; 102:1419–1423. [PubMed: 15668396]

9. West JB, Maurer CR. Designing optically tracked instruments for image-guided surgery. *IEEE Trans Med Imag.* 2004; 23, no. 5:533–545.
10. Ma B, Ellis RE. Analytic expressions for fiducial and surface target registration error. *Med Image Comput Comput-Assist Intervention (MICCAI)*. 2006; 4191:637–644.
11. Ma B, Moghari MH, Ellis RE, Abolmaesumi P. Estimation of optimal fiducial target registration error in the presence of heteroscedastic noise. *IEEE Trans Med Imag.* 2010; 29, no. 3:708–723.
12. Moghari MH, Abolmaesumi P. Distribution of target registration error for anisotropic and inhomogeneous fiducial localization error. *IEEE Trans Med Imag.* 2009; 28, no. 6:799–813.
13. Umeyama S. Least-squares estimation of transformation parameters between two point patterns. *IEEE Trans Pattern Anal.* 1991; 13, no. 4:376–380.
14. Fitzpatrick JM, West JB, Maurer CR. Predicting error in rigid-body point-based registration. *IEEE Trans Med Imag.* 1998; 17, no. 9:694–702.
15. Fitzpatrick JM, West JB. The distribution of target registration error in rigid-body point-based registration. *IEEE Trans Med Imag.* 2001; 20:917–927.
16. West JB, Fitzpatrick JM, Toms SA, Maurer CR, Maciunas RJ. Fiducial point placement and the accuracy of point-based, rigid body registration. *Neurosurg.* 2001; 48:810–817.
17. Arun KS, Huang TS, Blostein SD. Least-squares fitting of two 3D point sets. *IEEE Trans Pattern Anal.* 1987; 9, no. 5:698–700.
18. Davies RB, Hutton B. The effect of errors in the independent variables in linear regression. *Biometrika.* 1975; 62:383–391.
19. Van Huffel, S. *The Total Least Squares Problem: Computational Aspects and Analysis*. Philadelphia, PA, USA: SIAM; 1991.
20. Van Huffel, S.; Lemmerling, P., editors. *Total Least Squares and Errors-in-Variables Modeling*. Boston, MA, USA: Kluwer Academic; 2002.
21. Markovsky I, Van Huffel S. Overview of total least-squares methods. *Signal Process.* 2007; 87:2283–2302.
22. Gleser LJ. Estimation in a multivariate “errors in variable” regression model: Large sample results. *Ann Statist.* 1981; 9:24–44.
23. Yetik IS, Nehorai A. Performance bounds on image registration. *IEEE Trans Signal Process.* 2006; 54, no. 5:1737–1749.
24. Li J, Huang P. A comment on “performance bounds on image registration”. *IEEE Trans Signal Process.* 2009; 57, no. 6:2432–2433.
25. Koyama-Honda I, Ritchie K, Iion R, Murakoshi H, Kasai RS, Kusumi A. Fluorescence imaging for monitoring the colocalization of two single molecules in living cells. *Biophys J.* 2005; 88:2126–2136. [PubMed: 15596511]
26. Ober RJ, Ram S, Ward ES. Localization accuracy in single-molecule microscopy. *Biophys J.* 2004; 86:1185–1200. [PubMed: 14747353]
27. Ram S, Ward ES, Ober RJ. Beyond Rayleigh's criterion: a resolution measure with application to single-molecule microscopy. *PNAS.* 2006; 103:4457–4462. [PubMed: 16537357]
28. Ram S, Ward ES, Ober RJ. A stochastic analysis of distance estimation approaches in single molecule microscopy: Quantifying the resolution limits of photon-limited imaging systems. *Multidim Syst Signal Process.* 2012; 3:503–542.
29. Shen Z, Andersson SB. Bias and precision of the fluoroBrancoft algorithm for single particle localization in fluorescence microscopy. *IEEE Trans Signal Process.* 2011; 59, no. 8:4041–4046.
30. Wong Y, Lin Z, Ober RJ. Limit of the accuracy of parameter estimation for moving single molecules imaged by fluorescence microscopy. *IEEE Trans Signal Process.* 2011; 59, no. 3:895–911.
31. Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC. Green fluorescent protein as a marker for gene expression. *Sci.* 1994; 263:802–805.
32. Pierce DW, Hom-Booher N, Vale RD. Imaging individual green fluorescent proteins. *Nature.* 1997; 388:338. [PubMed: 9237750]

33. Schmidt T, Schütz GJ, Baumgartner W, Gruber HJ, Schindler H. Characterization of photophysics and mobility of single molecules in a fluid lipid membrane. *J Phys Chem -US*. 1995; 99:17662–17668.
34. Matei B, Meer P. Optimal rigid motion estimation and performance evaluation with bootstrap. *Proc IEEE Comput Vis Pattern Recogn*. 1999; 1:1339–1342.
35. Matei B, Meer P. A general method for errors-in-variables problems in computer vision. *Proc IEEE Comput Vis Pattern Recogn*. 2000; 2:2018–2021.
36. Chan NN, Mak TK. Heteroscedastic errors in a linear-functional relationship. *Biometrika*. 1984; 71:212–215.
37. Cohen EAK, Ober RJ. Image registration error analysis with applications in single molecule microscopy. *Proc I S Biomed Imag*. 2012:996–999.
38. Neyman J, Scott EL. Consistent estimates based on partially consistent observations. *Econometrica*. 1948; 19:1–32.
39. Morton R. Efficiency of estimating equations and the use of pivots. *Biometrika*. 1981; 68:227–233.
40. Foutz RV. On the unique consistent solution to the likelihood equations. *J Amer Stat Assoc*. 1977; 72:147–148.
41. Chan NN, Mak TK. Estimation in multivariate errors-in-variables models. *Linear Algebra Appl*. 1985; 70:197–207.
42. Ram S, Prabhat P, Ward ES, Ober RJ. Improved single particle localization accuracy with dual objective multifocal plane microscopy. *Opt Express*. 2009; 17:6881–6898. [PubMed: 19365515]
43. French AP, Mills S, Swarup R, Bennett MJ, Pridmore TP. Colocalization of fluorescent markers in confocal microscope images of plant cells. *Nat Protoc*. 2008; 3:619–628. [PubMed: 18388944]
44. Lachmanovich E, Shvartsman DE, Malka Y, Botvin C, Henis YI, Weiss AM. Co-localization analysis of complex formation among membrane proteins by computerized fluorescence microscopy: Application to immunofluorescence co-patching studies. *J Microsc*. 2003; 212, no. 2:122–131. [PubMed: 14629561]
45. Cohen EAK, Ober RJ. Measurement errors in fluorescence microscopy experiments. in *Conf Rec Asilomar C*. 2012:1602–1606.

## Biographies

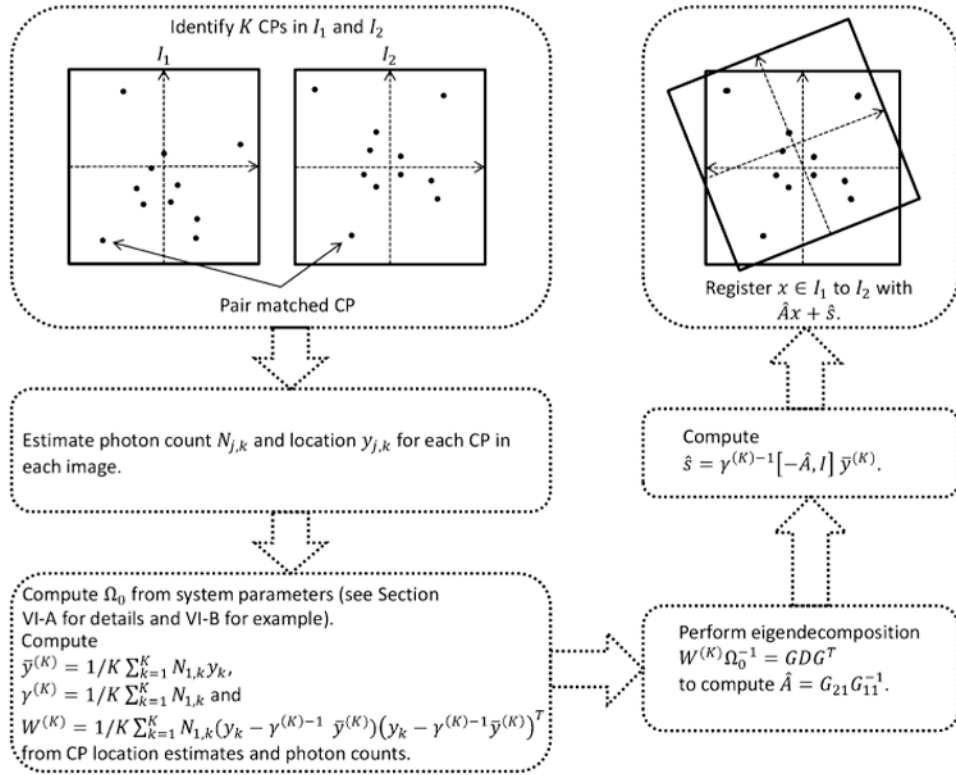


**E. A. K. Cohen** received the M.Phys. degree in mathematics and physics from the University of Warwick, Coventry, U.K., in 2005 and the Ph.D. degree in 2011 from Imperial College London, U.K., in statistics. He worked as a Signal Processing Scientist at QinetiQ, Malvern, U.K., before starting his Ph.D. in 2007. In 2011, he joined the University of Texas (UT), Dallas, in the Ward-Ober lab based at UT Southwestern Medical Center at Dallas as a Postdoctoral Research Scholar. He is currently a Lecturer in Statistics at Imperial College London. His research interests include multivariate signal and image processing, and their applications to the natural sciences.

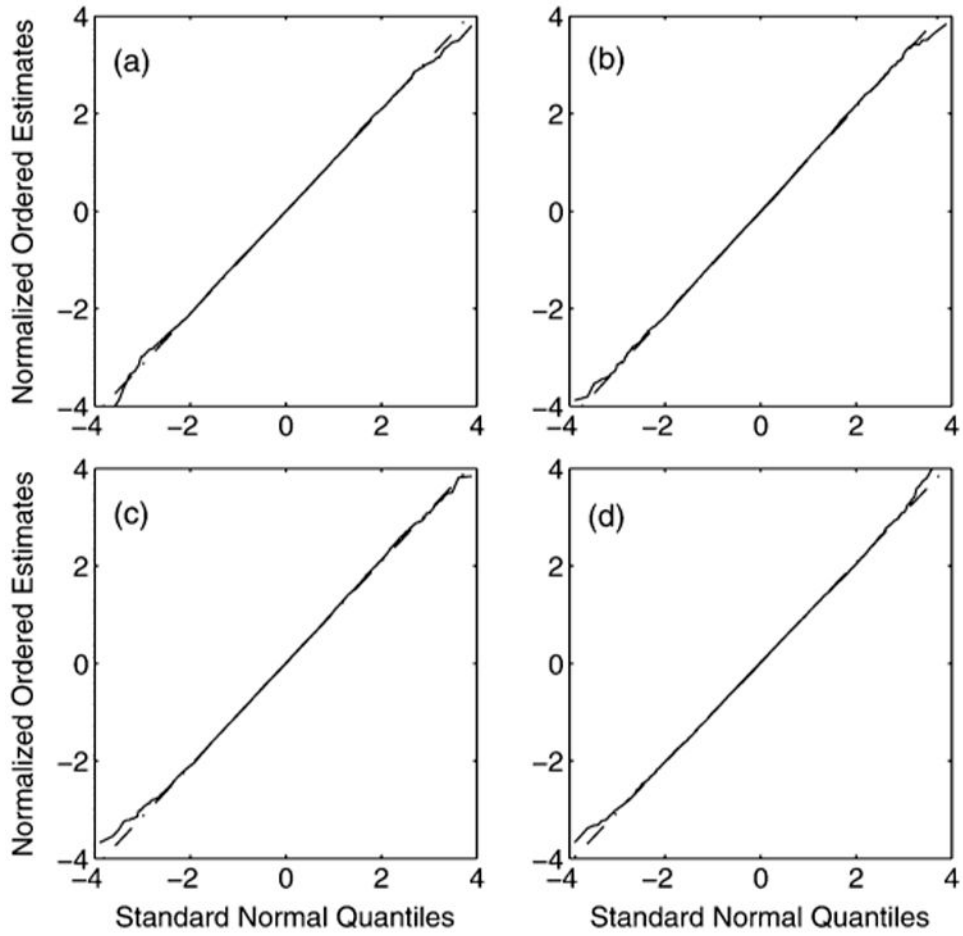


**R. J. Ober** (S'87-M'88-SM'95) received the Ph.D. degree in electrical engineering from the University of Cambridge, U.K., in 1987.

From 1987 to 1990, he was a Research Fellow at Girton College and the Engineering Department, University of Cambridge. In 1990, he joined the University of Texas at Dallas, Richardson, where he is currently a Professor with the Department of Electrical Engineering and the Department of Bioengineering. He is also Adjunct Professor at the University of Texas Southwestern Medical Center, Dallas. His research interests include signal and imaging processing, and their applications to biological problems, fluorescence microscopy and biosensors.

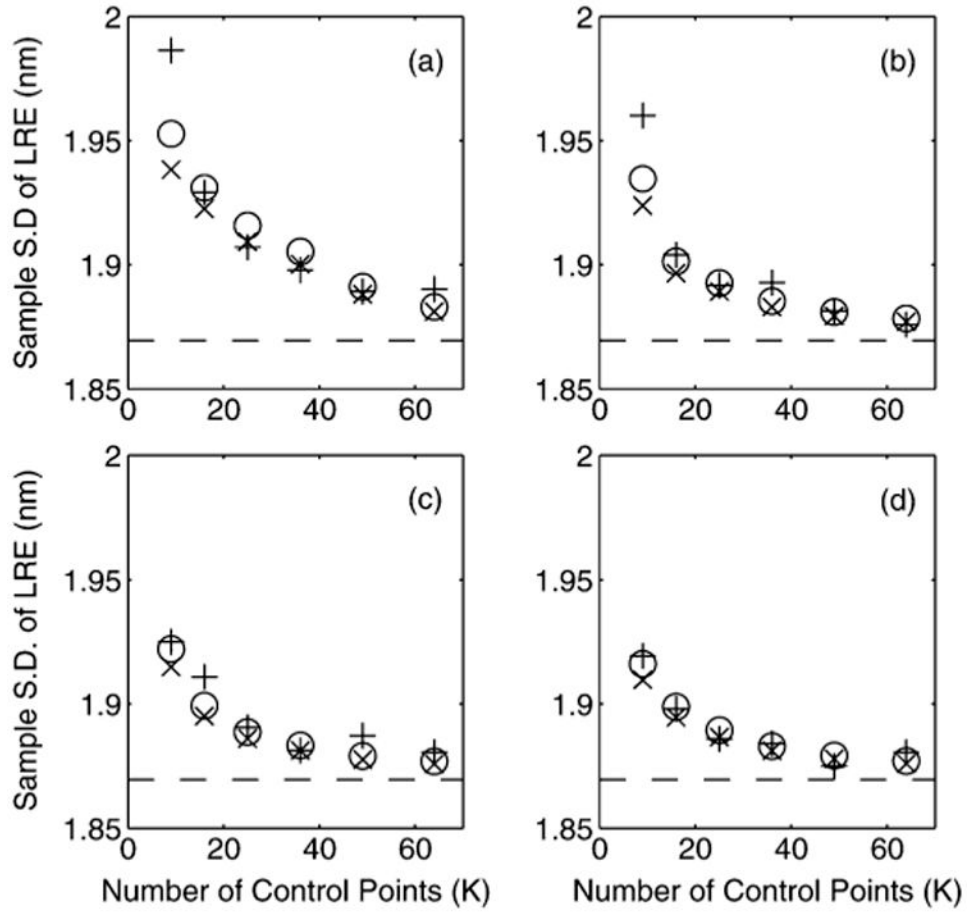


**Fig. 1.** Flow chart summarizing the key steps in registering a pair of fluorescence microscopy images.

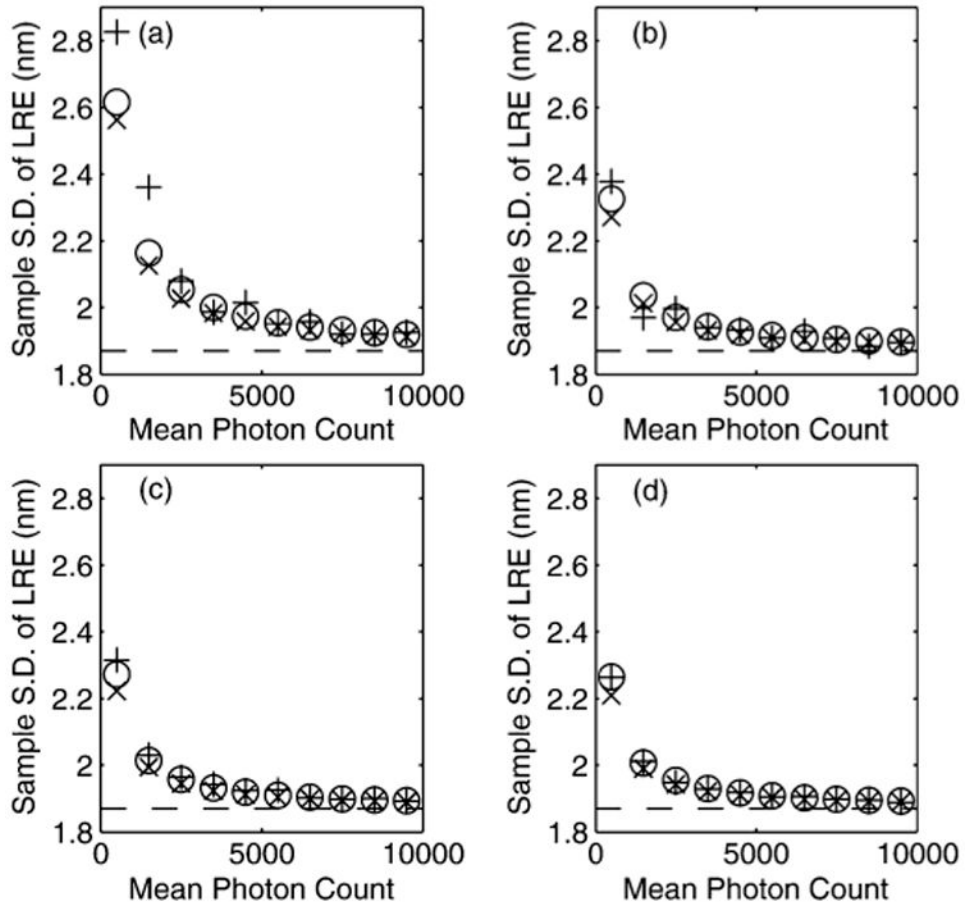


**Fig. 2.** Q-Q plot for the first element of the TRE vector with normally distributed CP measurement errors, for (a)  $K = 6$ , (b) 10, (c) 15, (d) 20 CPs uniformly distributed in the image space. The (mostly obscured) dashed straight line marks the line of perfect fit.

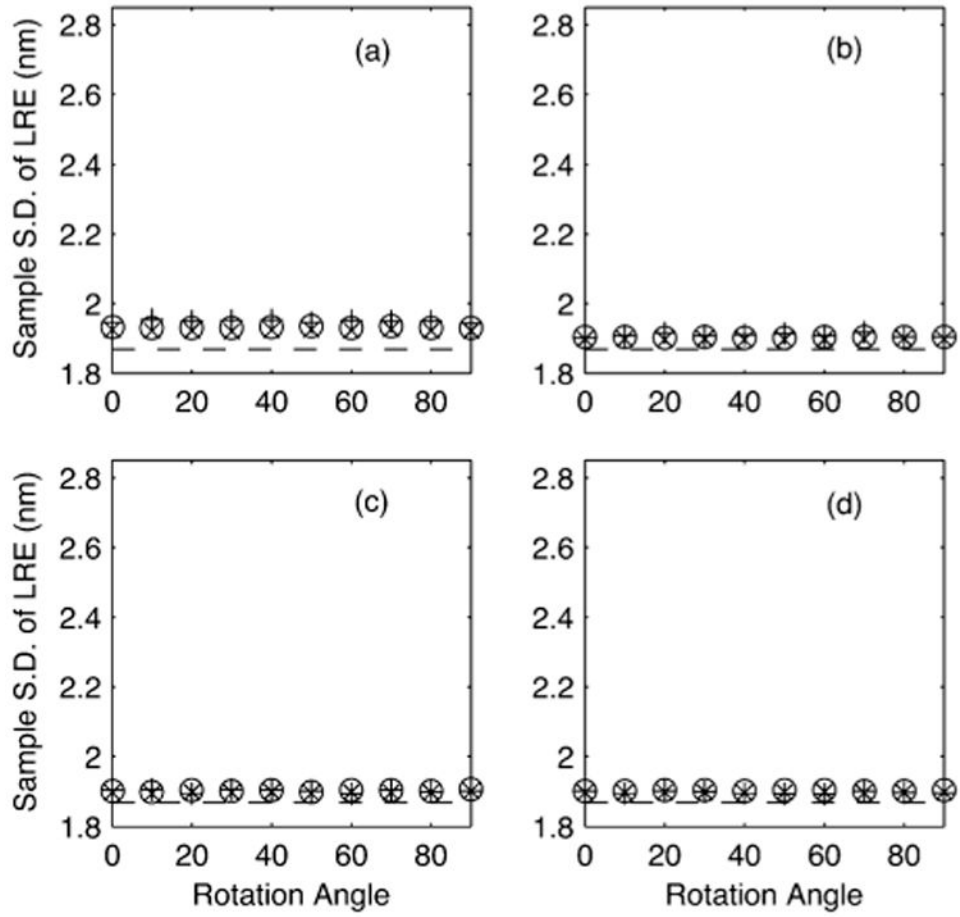




**Fig. 3.** The sample standard deviation of  $\ell_F$  (from 100000 simulations) in object space dimensions is plotted as a function of the number of CPs  $K$  for (a) normally distributed, (b) uniformly distributed, (c) elliptical and (d) grid CP configurations (see Section VII-B for more details). The '+' represents the sample standard deviation of the LRE (in units nanometers). The circles represent the standard deviation as predicted by Corollary V.2. The 'x' shows the standard deviation as predicted with (18). The dashed line marks the theoretical bound  $\sigma_F$ .

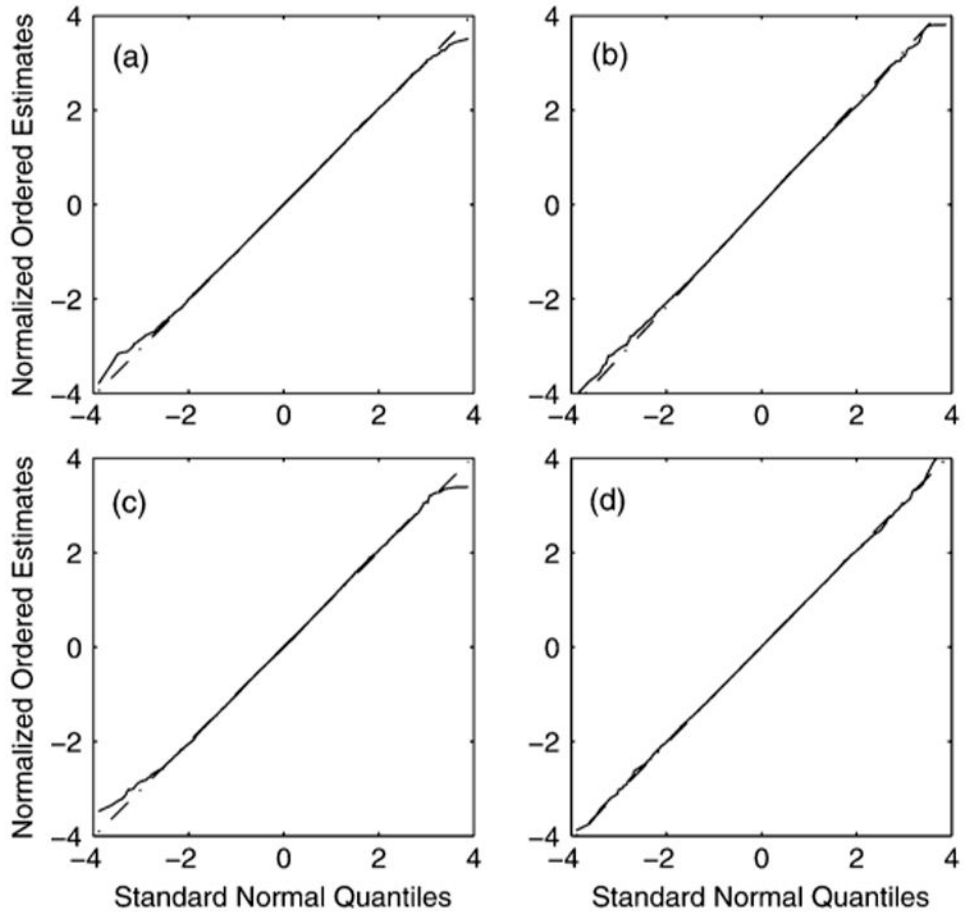


**Fig. 4.** The sample standard deviation of  $\ell_F$  (from 100000 simulations) in object space dimensions is plotted as a function of  $N_1$ , the mean photon count for the CPs in  $I_1$ , for a constant  $K = 16$  with (a) normally distributed, (b) uniformly distributed, (c) elliptical and (d) grid CP configurations (see Section VII-B for more details). The '+' represents the sample standard deviation of the LRE (in units nanometers). The circles represent the standard deviation as predicted by Corollary V.2. The 'x' shows the standard deviation as predicted with (18). The dashed line marks the theoretical bound  $\sigma_F$ .



**Fig. 5.**

The sample standard deviation of  $\ell_F$  (from 100000 simulations) in object space dimensions is plotted as a function of rotation angle  $\phi$  (in degrees) when  $A = R$  with  $R$  of form (21), for a constant  $K = 16$  with (a) normally distributed, (b) uniformly distributed, (c) elliptical and (d) grid CP configurations (see Section VII-B for more details). The '+' represents the sample standard deviation of the LRE (in units nanometers). The circles represent the standard deviation as predicted by Corollary V.2. The 'x' shows the standard deviation as predicted with (18). The dashed line marks the theoretical bound  $\sigma_F$ .



**Fig. 6.** Q-Q plot for the first element of the TRE vector with uniformly distributed CP measurement errors, for (a)  $K = 6$ , (b) 10, (c) 15, (d) 20 CPs uniformly distributed in the image space. The (mostly obscured) dashed straight line marks the line of perfect fit.

**Table I**

Sample standard deviations of TRE in the  $x$ - and  $y$ -directions from 599 registration experiments (units: mm), each row is for one of 27 different coordinate points. ‘WCGLS’ columns are using the New weighted covariance generalized least squares presented in This paper. ‘GLS’ columns are using traditional generalized least Squares that assumes homoscedastic measurement errors. ‘GAIN’ Columns give the percentage improvement in the WCGLS method Over the GLS method

WCGLS (x) - mm	GLS (x) - mm	Gain (%)	WCGLS (y) - mm	GLS (y) - mm	Gain (%)
4.142	4.283	3.399	4.003	4.342	8.478
3.942	4.039	2.446	3.762	4.095	8.873
4.030	4.736	17.517	4.111	4.531	10.208
3.657	4.238	15.880	3.640	4.001	9.926
2.972	3.073	3.393	2.659	2.906	9.292
2.709	2.896	6.916	2.361	2.581	9.331
3.124	3.187	2.026	2.846	3.156	10.886
3.913	4.082	4.326	3.801	4.256	11.954
2.338	2.436	4.206	1.842	2.021	9.717
3.436	3.844	11.897	3.194	3.573	11.851
3.743	4.207	12.382	3.555	3.984	12.055
4.279	4.684	9.461	4.355	4.896	12.426
3.363	3.660	8.850	3.302	3.692	11.811
2.865	3.053	6.555	2.441	2.693	10.343
2.472	2.566	3.796	1.976	2.145	8.554
3.623	3.865	6.699	3.300	3.657	10.842
3.404	3.772	10.812	3.403	3.789	11.352
2.757	2.853	3.484	2.364	2.536	7.276
2.896	3.146	8.623	2.733	2.974	8.800
3.154	3.274	3.800	2.859	3.053	6.791
4.674	4.857	3.929	4.401	4.840	9.985
3.449	3.831	11.060	3.462	3.750	8.318
3.282	3.537	7.774	3.162	3.390	7.187
3.282	3.465	5.592	3.085	3.292	6.717

WCGLS (x) - mm	GLS (x) - mm	Gain (%)	WCGLS (y) - mm	GLS (y) - mm	Gain (%)
3.687	3.985	8.074	3.572	3.870	8.349
3.987	4.043	1.401	3.799	4.201	10.595
2.590	2.832	9.359	2.183	2.412	10.514

**Table II****Abbreviations**

CP(s)	control point(s)
TRE	target registration error - Definition II. 1
EIV	errors-in- variables - (1)
TLS	total least squares
ML	maximum likelihood
HEIV	heteroscedastic errors-in-variables
LRE	localization registration error - Definition II.2
GLS	generalized least squares - Definition III. 1
OLS	ordinary least squares - Definition III.1
SPD	symmetric positive definite
WCGLS	weighted covariance generalized least squares - Section III-A

**Table III**  
**Operators and Functions**

$\text{cov}\{\cdot\}$	covariance matrix of a random vector
$N_p(\mu, \Sigma)$	p-dimensional multivariate normal distribution with mean $\mu$ and covariance matrix $\Sigma$
$\equiv$	equal in distribution
$\mathcal{L}(\cdot)$	likelihood function
$p(\cdot)$	probability density function
$\text{vec}(\cdot)$	$\text{vec}(Z)$ of a $r \times a$ matrix $Z = [z_1, \dots, z_s]$ is the $rs \times 1$ vector $\begin{bmatrix} z_1^T \\ \vdots \\ z_s^T \end{bmatrix}$
$E\{\cdot\}$	expectation operator
$\otimes$	Kronecker product
$\delta_{ij}$	$\delta_{ij}$ equals zero for $i \neq j$ and equals one for $i = j$ .



**Table IV**  
**Key Notation**

Notation	Type	Definition
$d$	$d \in \{2,3\}$	dimension of image space
$\mathcal{J}$	subset of $\mathbb{R}^d$	image space $j, j = 1,2$
$A$	$\mathbb{R}^{d \times d}$	matrix valued affine transformation parameter
$s$	$\mathbb{R}^d$	vector valued affine transformation parameter
$K$	$\mathbb{N}$	number of CPs
$x_{j,k}$	$\mathbb{R}^d$	true location of $k$ th CP ( $k = 1, \dots, K$ ) in $\mathcal{J}$
$\epsilon_{j,k}$	$\mathbb{R}^d$	random measurement error associated with localizing $k$ th CP in $\mathcal{J}$
$y_{j,k}$	$\mathbb{R}^d$	measured location of $k$ th CP in $\mathcal{J}$ , $y_{j,k} = x_{j,k} + \epsilon_{j,k}$
$\Omega_{j,k}$	$\mathbb{R}^{d \times d}$	$\Omega_{j,k} = \text{cov}\{\epsilon_{j,k}\}$
$X_j$	$\mathbb{R}^{d \times K}$	$X_j = [x_{j,1}, \dots, x_{j,K}]$
$\mathcal{E}_j$	$\mathbb{R}^{d \times K}$	$\mathcal{E}_j = [\epsilon_{j,1}, \dots, \epsilon_{j,K}]$
$Y_j$	$\mathbb{R}^{d \times K}$	$Y_j = [y_{j,1}, \dots, y_{j,K}]$
$X$	$\mathbb{R}^{2d \times K}$	$X = [X_1^T, X_2^T]^T$
$\mathcal{E}$	$\mathbb{R}^{2d \times K}$	$\mathcal{E} = [\mathcal{E}_1^T, \mathcal{E}_2^T]^T$
$Y$	$\mathbb{R}^{2d \times K}$	$Y = [Y_1^T, Y_2^T]^T$
$\Lambda$	$\mathbb{R}^{2d \times d}$	$\Lambda = [I_{2d}, A^T]^T$
$\mathbf{0}$	$\mathbb{R}^d$	$\mathbf{0} = [0, 0]^T$ or $[0, 0, 0]^T$
$a$	$\mathbb{R}^{2d}$	$a = [\mathbf{0}^T, s^T]^T$
$\mathbf{1}_K$	$\mathbb{R}^K$	$\mathbf{1} = [1, 1, \dots, 1]^T$
$\epsilon_k$	$\mathbb{R}^{2d}$	$\epsilon_k = [\epsilon_{1,k}^T, \epsilon_{2,k}^T]^T$
$\Omega_k$	$\mathbb{R}^{2d \times 2d}$	$\Omega_k = \text{cov}\{\epsilon_k\}$
$\tau(\cdot)$	$\mathbb{R}^d \mapsto \mathbb{R}^d$	<b>TRE</b>
$x_{j,F}$	$\mathbb{R}^d$	true location of a feature in $\mathcal{J}$
$\epsilon_{j,F}$	$\mathbb{R}^d$	random measurement error associated with localizing a feature in $\mathcal{J}$
$y_{j,F}$	$\mathbb{R}^d$	measured location of a feature in $\mathcal{J}$ , $y_{j,F} = x_{j,F} + \epsilon_{j,F}$
$\Omega_{1,F}$	$\mathbb{R}^{d \times d}$	$\Omega_F = \text{cov}\{\epsilon_{1,F}\}$
$\hat{A}$	$\mathbb{R}^{d \times d}$	estimator of $A$
$\hat{s}$	$\mathbb{R}^d$	estimator of $s$
$\mathcal{L}_F$	$\mathbb{R}^d$	<b>LRE</b> for a feature
$\Delta A$	$\mathbb{R}^{d \times d}$	$\Delta A = \hat{A} - A$

Notation	Type	Definition
$\Delta s$	$\mathbb{R}^d$	$\Delta s = \hat{s} - s$
$\Omega_\tau$	$\mathbb{R}^{d \times d}$	$\Omega_\tau = \text{cov}\{\tau\}$
$\Omega_\ell$	$\mathbb{R}^{d \times d}$	$\Omega_\ell = \text{cov}\{\ell_F\}$
$y_k$	$\mathbb{R}^{2d}$	$y_k = [y_{1,k}^T, y_{2,k}^T]^T$
$U$	$\mathbb{R}^{d \times 2d}$	$U = [-A, I_d]$
$\Phi_k$	$\mathbb{R}^{d \times d}$	$\Phi_k = U\Omega_k U^T$
$\beta$	$\mathbb{R}^{(d+d^2)}$	$\beta = [s^T, \text{vec}(A)^T]^T$
$\eta_k$	$\mathbb{R}^+$	scalar and matrix valued quantities when we have the property $\Omega_k = \eta_k \Omega_0$
$\Omega_0$	$\mathbb{R}^{2d \times 2d}$	
$\gamma^{(K)}$	$\mathbb{R}^+$	$\gamma^{(K)} = (1/K) \sum_{k=1}^K \eta_k^{-1}$
$\gamma$		$\gamma = \lim_{K \rightarrow \infty} \gamma^{(K)}$
$\bar{y}^{(K)}$	$\mathbb{R}^{2d}$	$\bar{y}^{(K)} = (1/K) \sum_{k=1}^K \eta_k^{-1} y_k$
$\bar{y}$		$\bar{y} = \lim_{K \rightarrow \infty} \bar{y}^{(K)}$
$W^{(K)}$	$\mathbb{R}^{2d \times 2d}$	$W^{(K)} = (1/K) \sum_{k=1}^K \eta_k^{-1} \times (y_k - \gamma^{(K)-1} \bar{y}^{(K)}) \times (y_k - \gamma^{(K)-1} \bar{y}^{(K)})^T$
$W$		$W = \lim_{K \rightarrow \infty} W^{(K)}$
$\bar{x}_1^{(K)}$	$\mathbb{R}^d$	$\bar{x}_1^{(K)} = (1/K) \sum_{k=1}^K \eta_k^{-1} x_{1,k}$
$\Xi^{(K)}$	$\mathbb{R}^{d \times d}$	$\Xi^{(K)} = (1/K) \sum_{k=1}^K \eta_k^{-1} x_{1,k} x_{1,k}^T$
$\Psi^{(K)}$	$\mathbb{R}^{d \times d}$	$\Psi^{(K)} = \Xi^{(K)} - \gamma^{(K)-1} \bar{x}_1^{(K)} \bar{x}_1^{(K)T}$
$\Psi$		$\Psi = \lim_{K \rightarrow \infty} \Psi^{(K)}$
$\Theta^{(K)}$	$\mathbb{R}^{d \times d}$	$\Theta^{(K)} = \Psi^{(K)-1} (\Lambda^T \Omega_0^{-1} \Lambda)^{-1} \Psi^{(K)-1} + \Psi^{(K)-1}$
$\Theta$		$\Theta = \lim_{K \rightarrow \infty} \Theta^{(K)}$
$\theta$	$\mathbb{R}$	$\theta = \bar{x}_1^T \Theta \bar{x}_1$
$\Phi_0$	$\mathbb{R}^{d \times d}$	$\Phi_0 = U\Omega_0 U^T$

**Table V**  
**Fluorescence Microscopy Notation**

Notation	Type	Definition
$N_{j,k}$	$\mathbb{N}$	photon count associated with $k$ th CP in $\mathcal{A}_j$
$N_j$	$\mathbb{Z}^+$	mean photon count for CPs in $\mathcal{A}_j$
$\mathcal{M}_j$	$\mathbb{R}^+$	system magnification for $\mathcal{A}_j$
$\lambda_{j,em}$	$\mathbb{R}^+$	photon wavelength associated with $\mathcal{A}_j$
$\zeta_j$	$\mathbb{R}^+$	known function of the point spread function, photon wavelength and numerical aperture for $\mathcal{A}_j$
$\kappa^2$	$\mathbb{R}^+$	$x/y$ -direction variance of randomly positioned CPs
$R$	$\mathbb{R}^{d \times d}$	a rotation matrix
$\zeta$	$\mathbb{R}^+$	scaling constant for $A$ of the form $\zeta R$
$\bar{\sigma}_j^2$	$\mathbb{R}^+$	$\bar{\sigma}_j^2 = \mathcal{M}_j^2 \zeta_j / \bar{N}_j$
$r_F$	$\mathbb{R}^+$	radial distance of single molecule from origin of $\mathcal{A}_j$