# Direct Estimation of Homogeneous Vectors: An Ill-Solved Problem in Computer Vision

Matthew Harker and Paul O'Leary

Institute for Automation, University of Leoben, A-8700 Leoben, Austria
matthew.harker@stud.unileoben.ac.at
http://automation.unileoben.ac.at

**Abstract.** Computer Vision theory is firmly rooted in Projective Geometry, whereby geometric objects can be effectively modeled by homogeneous vectors. We begin from Gauss's 200 year old theorem of least squares to derive a generic algorithm for the direct estimation of homogeneous vectors. We uncover the common link of previous methods, showing that direct estimation is not an ill-conditioned problem as is the popular belief, but has merely been an ill-solved problem. Results show improvements in goodness-of-fit and numerical stability, and demonstrate that "data normalization" is unnecessary for a well-founded algorithm.

## 1 Introduction

Geometric objects which can be modeled by homogeneous vectors range from implicit curves and surfaces, to the fundamental matrix describing epipolar geometry, and projective transformations such as camera matrices and homographies. Metric Vision tasks have real-time constraints, so it is critical to have fast and robust techniques for the estimation of homogeneous vectors. The basis of this paper is a plethora of seemingly unrelated techniques for the direct (i.e. non-iterative) estimation of homogeneous vectors. Direct estimation techniques have several advantages, namely, that the minimization yields the global minimum of the cost function, and they are computed in a finite number of steps. Three of the most commonly encountered methods for direct estimation are:

1. **Normalization:** The algorithm proceeds by normalizing the data, followed by minimization of the algebraic error subject to a unit norm constraint [1]. While the method has the advantage of simplicity, Hartley notes that without normalization, the algorithm is guaranteed to perform extremely poorly, which indicates the algorithm is poorly founded, not ill-conditioned. The unit norm constraint is typically justified by the homogeneity of the vector; however, we show that it is rarely a mathematically justifiable constraint. Specifically, for linear geometric models, the algorithm returns meaningless results. An algorithm that cannot handle linear models clearly cannot suffice for a generic methodology.
2. **Invariant Fitting:** This entails the least-squares minimization of algebraic error subject to a geometrically invariant constraint. Invariant fitting was originally proposed by Bookstein [2] to fit conics independent of the chosen

coordinate frame, and was later adapted to fundamental matrix estimation by Torr [3]. Other variations include fitting conics of specific types [4]. At this point in time, algorithms are limited to these applications, since viable invariant constraints must be quadratic [5]. This too corresponds to minimizing an algebraic distance, however, the results differ from the normalization scheme because a different constraint is imposed.

3. **Gradient-Weighted Fitting:** Sampson [6] proposed that the so-called algebraic distance weighted by its gradient would provide an improved metric for fitting conics, although his approach to attempt to minimize said cost function was iterative. Taubin [7] showed that an approximation to the minimum of the cost function can be found directly by generalized eigenvectors, which is often referred to as Gradient-Weighted Fitting. Here we show that while it has the potential to provide a better approximation, the algorithm is fundamentally unstable from a numerical point of view; in fact, it is far too unstable to yield useful results in Computer Vision applications (i.e. with pixel coordinates). Introducing data normalization alleviates, but does not rectify this instability.

In the present work, we derive a generic algorithm for the direct estimation of homogeneous vectors which (i) has the goodness-of-fit properties of gradient weighted fitting, however, with immeasurably improved stability, (ii) is largely invariant to the choice of coordinate frame and (iii) circumvents the need for data normalization. Simply put, it amalgamates the desirable properties of the most relied-upon techniques for the direct estimation of homogeneous vectors.

## 2    Gauss's Theorem of Least Squares

We hearken back to a two hundred year old theorem, Gauss's theorem of least-squares [8], which is central to estimation in the presence of uncertainty. Sadly, we find that it is largely misused, and the original theorem all but forgotten, save in a handful of Numerical Analysis literature. Gauss proposed various models for errors in measurements; the most fruitful was exponential-based, providing a realistic model that can be treated analytically. A measurement error is modeled as a random $n$-vector, $\delta$, which behaves according to the probability distribution

$$P(\delta) = ((2\pi)^n |\Lambda|)^{-\frac{1}{2}} \exp -\frac{1}{2} \delta^{\mathrm{T}} \Lambda^{-1} \delta, \tag{1}$$

where $\Lambda$ is the $n \times n$ covariance matrix [9,10]. This has come to be known as a Gaussian distribution.

**Least Squares as a Maximum Likelihood Criterion.** Gauss begins with a set of observations (i.e. measurements), which should conform to a linear model,

$$\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} e_1 \\ \vdots \\ e_m \end{bmatrix}. \tag{2}$$

Each error, $e_i$, is assumed to be an independent random variable that follows the Gaussian distribution, $\psi(e_i) = h\pi^{-\frac{1}{2}} \exp -h^2 e_i^2$, where $h$ is a positive constant; this assumes that all errors are mean-free and have the same variance. The function $\Omega$ is proposed, $\Omega = \prod_{i=1}^{n} \psi(e_i)$, the motivation being that for a normal distribution, the smallest error is the most probable: the function $\Omega$ should therefore be maximized. This function gives rise to the principle of maximum likelihood. Under the above assumptions, we have,

$$\Omega = h^n \pi^{-\frac{1}{2}n} \exp -h^2 \left( e_1^2 + e_2^2 + \ldots + e_n^2 \right). \tag{3}$$

Hence, to maximize the likelihood, $\Omega$, we must minimize,

$$\epsilon \triangleq e_1^2 + e_2^2 + \ldots + e_n^2 = \sum_{i=1}^{n} e_i^2, \tag{4}$$

the sum of squared errors. Gauss later proved that if the variances of the errors are all scaled to unity, then the least squares solution is such that the estimation errors have minimal variance. Summarizing the postulates:

1. Errors in measurements behave according to Gaussian distributions.
2. The errors are mean-free and scaled such that they have a unit variance.

The significance of the two postulates is clear. If the errors are not mean-free and normally distributed with equal variances, the least-squares solution is no longer a maximizer of the likelihood function $\Omega$.

## 3   Linear Models

Linear models of geometric objects in two, three, and $n$ dimensions, are respectively lines, planes, and hyperplanes. The following theory applies generally to these models, but for simplicity and visualization, we specifically address lines in the plane. The homogeneous equation of a line in the plane is given as,

$$\mathbf{p}^\mathsf{T}\mathbf{z} = ax + by + c = 0, \tag{5}$$

where

$$\mathbf{p} = \begin{bmatrix} x\ y\ 1 \end{bmatrix}^\mathsf{T} \qquad \text{and} \qquad \mathbf{z} = \begin{bmatrix} a\ b\ c \end{bmatrix}^\mathsf{T}. \tag{6}$$

In practice, we measure points $(\hat{x}, \hat{y})$, which do not lay on the line, but deviate by some error which we model as the random variables $(\delta_x, \delta_y)$, such that

$$(\hat{x}, \hat{y}) = (x + \delta_x, y + \delta_y). \tag{7}$$

For mathematical convenience, we model the random coordinate pair $\delta = (\delta_x, \delta_y)$ as mean-free and correlated according to the Gaussian distribution,

$$P(\delta_x, \delta_y) = \frac{1}{2\pi} \left( \left| \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{xy} & \sigma_{yy} \end{bmatrix} \right| \right)^{-\frac{1}{2}} \exp -\frac{1}{2} \begin{bmatrix} \delta_x & \delta_y \end{bmatrix} \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{xy} & \sigma_{yy} \end{bmatrix}^{-1} \begin{bmatrix} \delta_x \\ \delta_y \end{bmatrix}, \tag{8}$$

defined by the covariance matrix, $\Lambda$. Since the ideal point $(x, y)$ fits the model, we may write,

$$a(\hat{x} - \delta_x) + b(\hat{y} - \delta_y) + c = 0. \tag{9}$$

Rearranging yields,

$$a\hat{x} + b\hat{y} + c = a\delta_x + b\delta_y. \tag{10}$$

The left hand side is the familiar algebraic residual error associated with a point and a line, which we denote as, $r_{\mathrm{a}}$, such that

$$r_{\mathrm{a}}(\hat{x}, \hat{y}, \mathbf{z}) \triangleq a\hat{x} + b\hat{y} + c. \tag{11}$$

The right hand side tells us how the error in the algebraic residual behaves according to the random errors in the point coordinates $(\delta_x, \delta_y)$; we denote the right hand side as $r_{\mathrm{s}}$, the stochastic form of the residual, such that

$$r_{\mathrm{s}}(\delta_x, \delta_y, \mathbf{z}) \triangleq a\delta_x + b\delta_y. \tag{12}$$

Since we have assumed a mathematical model for the error $(\delta_x, \delta_y)$, we may analytically compute the expected value and variance of the error [9],

$$E[r] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(\delta_x, \delta_y) r_{\mathrm{s}}(\delta_x, \delta_y, \mathbf{z}) \mathrm{d}\delta_x \mathrm{d}\delta_y = 0, \tag{13}$$

and

$$\mathrm{Var}\,(r) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(\delta_x, \delta_y)(r_{\mathrm{s}}(\delta_x, \delta_y, \mathbf{z}) - E[r])^2 \mathrm{d}\delta_x \mathrm{d}\delta_y \tag{14}$$

$$= a^2 \sigma_{xx} + 2ab\sigma_{xy} + b^2 \sigma_{yy}. \tag{15}$$

There are a few important points to note:

1. The stochastic form of the residual is independent of the constant term, $c$, hence so are the mean and variance. In the case of some linearized models, such as homographies or camera matrices, there are multiple constant terms.
2. The variance is clearly a quadratic form in the statistically dependent coefficients, which will always be the case since variance is quadratic.

In light of these facts, we first partition the vector of unknowns into its statistically dependent and independent terms as

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}_A \\ \mathbf{z}_B \end{bmatrix} \quad \text{where for the line,} \quad \mathbf{z}_A = \begin{bmatrix} a \\ b \end{bmatrix} \quad \text{and} \quad \mathbf{z}_B = c. \tag{16}$$

As will be seen, this partitioning is critical, although it is largely ignored in the literature. We may therefore write the variance of the error, $r$, as the quadratic form, $\mathrm{Var}\,(r) = \mathbf{z}_A^{\mathrm{T}} \Lambda \mathbf{z}_A$. Now, if we are to properly implement a least-squares solution, each error should be weighted such that it has unit variance. We hence write the $i^{\mathrm{th}}$ error as,

$$\breve{e}_i = \frac{r_i}{\sqrt{\mathrm{Var}\,(r_i)}} = \frac{\mathbf{p}_i^{\mathrm{T}} \mathbf{z}}{\sqrt{\mathbf{z}_A^{\mathrm{T}} \Lambda_i \mathbf{z}_A}}, \tag{17}$$

where the notation $\check{e}$, is to stress the fact that $\check{e} \neq e$, but is however an error with unit variance[1]. We may now use the least-squares criterion; the cost function is

$$\epsilon(\mathbf{z}_A, \mathbf{z}_B) = \sum_{i=1}^{n} \check{e}_i^2 = \sum_{i=1}^{n} \frac{\left(\mathbf{p}_i^{\mathrm{T}} \mathbf{z}\right)^2}{\mathbf{z}_A^{\mathrm{T}} \Lambda_i \mathbf{z}_A} \tag{18}$$

The function $\epsilon$ is now in appropriate form such that minimizing $\epsilon$ corresponds to maximizing $\Omega$.

**On the Unit Norm Constraint.** For argument's sake, if we assume each error has the identity matrix as covariance matrix, i.e., $\Lambda = \mathsf{I}$, then the variance of each residual is $\mathrm{Var}\,(r) = a^2 + b^2$. The variance is clearly a function of the line direction; to ensure that the variance of the errors is independent of the line direction, we impose the constraint, $a^2 + b^2 = \alpha^2 = $ constant. The resulting variances of the residuals are constant with respect to the direction of the line. This uncovers the first fault in imposing a unit norm constraint on a homogeneous vector. For the line, this would mean $a^2 + b^2 + c^2 = 1$. The resulting variance of each residual would be, $\mathrm{Var}\,(r) = 1 - c^2$; since $c$ is the scaled distance of the line to the origin, this means that the variance of the error is functionally dependent on the position of the point in the plane, which is preposterous.

## 3.1   Minimizing the Least-Squares Cost Function

For convenience of manipulation, the cost function $\epsilon$ in Equation (18) can be written in matrix form, namely, as the squared 2-norm of a residual vector, i.e.,

$$\epsilon(\mathbf{z}_A, \mathbf{z}_B) = \|\mathsf{W}_A \mathsf{D}_A \mathbf{z}_A + \mathsf{W}_A \mathsf{D}_B \mathbf{z}_B\|_2^2, \tag{19}$$

where

$$\mathsf{W}_A = \mathrm{diag}\left(\left(\mathbf{z}_A^{\mathrm{T}} \Lambda_1 \mathbf{z}_A\right)^{-\frac{1}{2}}, \ldots, \left(\mathbf{z}_A^{\mathrm{T}} \Lambda_n \mathbf{z}_A\right)^{-\frac{1}{2}}\right), \tag{20}$$

and the subscript indicates the functional dependence $\mathsf{W}_A = \mathsf{W}_A(\mathbf{z}_A)$. For the case of the line, we have

$$\mathsf{D}_A = \begin{bmatrix} \hat{x}_1 & \hat{y}_1 \\ \vdots & \vdots \\ \hat{x}_n & \hat{y}_n \end{bmatrix} \quad \text{and} \quad \mathsf{D}_B = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}. \tag{21}$$

The residual vector is linear in $\mathbf{z}_B$, hence, $\epsilon$ is minimal when,

$$\mathbf{z}_B = -\left(\mathsf{W}_A \mathsf{D}_B\right)^+ \mathsf{W}_A \mathsf{D}_A \mathbf{z}_A, \tag{22}$$

where $\left(\mathsf{W}_A \mathsf{D}_B\right)^+$ is the Moore-Penrose pseudo-inverse [11]. Substituting Equation (22) into the cost function, it takes the general form,

---

[1] This variance weighted error is also known as the Mahalanobis distance from the point to the line, plane, or hyperplane.

$$\epsilon(\mathbf{z}_A) = \left\| \mathsf{W}_A \mathsf{D}_A \mathbf{z}_A - \mathsf{W}_A \mathsf{D}_B \left( \mathsf{W}_A \mathsf{D}_B \right)^+ \mathsf{W}_A \mathsf{D}_A \mathbf{z}_A \right\|_2^2. \tag{23}$$

Note the functional dependence of the cost function; this reduction corresponds to an oblique form of the Eckart-Young projection onto the constrained portion of the residual vector [12]. The reduced cost function corresponds to the Variable Projection (or VARPRO) method, whereby proof that $\mathbf{z}_A^*$ attaining the global minimum of $\epsilon(\mathbf{z}_A)$ with $\mathbf{z}_B$ given as in Equation (22) is equivalent to finding the global minimum of the function $\epsilon(\mathbf{z}_A, \mathbf{z}_B)$ can be found in [13]. Minimizing the cost function depends on the nature of the covariance matrices; we enumerate the three special cases as follows:

**Identical Covariance Matrices.** If all covariance matrices are identical, then the weighting matrix $\mathsf{W}_A$ can be written as $\mathsf{W}_A = \left( \mathbf{z}_A^{\mathrm{T}} \mathsf{\Lambda} \mathbf{z}_A \right)^{-\frac{1}{2}} \mathsf{I}$. The cost function simplifies to,

$$\epsilon = \frac{\left\| \mathsf{D}_A \mathbf{z}_A - \mathsf{D}_B \mathsf{D}_B^+ \mathsf{D}_A \mathbf{z}_A \right\|_2^2}{\mathbf{z}_A^{\mathrm{T}} \mathsf{\Lambda} \mathbf{z}_A} = \frac{\mathbf{z}_A^{\mathrm{T}} \mathsf{D}_A^{\mathrm{T}} \left( \mathsf{I} - \mathsf{D}_B \mathsf{D}_B^+ \right) \mathsf{D}_A \mathbf{z}_A}{\mathbf{z}_A^{\mathrm{T}} \mathsf{\Lambda} \mathbf{z}_A}, \tag{24}$$

If we define the matrix $\mathsf{S} \triangleq \mathsf{D}_A^{\mathrm{T}} \left( \mathsf{I} - \mathsf{D}_B \mathsf{D}_B^+ \right) \mathsf{D}_A$, then the cost function takes the form

$$\epsilon = \frac{\mathbf{z}_A^{\mathrm{T}} \mathsf{S} \mathbf{z}_A}{\mathbf{z}_A^{\mathrm{T}} \mathsf{\Lambda} \mathbf{z}_A}, \tag{25}$$

which is known as the Rayleigh quotient [11]. Indeed, we are interested in the extrema of this quotient, which in turn will yield the global minimum of the cost function. The extrema are, in fact, the eigenvalues and eigenvectors of the corresponding generalized eigenvalue problem,

$$\left( \mathsf{S} - \epsilon \mathsf{\Lambda} \right) \mathbf{z}_A = \mathbf{0}. \tag{26}$$

The global minimum is attained with the generalized eigenvector, $\mathbf{z}_A^*$, corresponding to the minimum eigenvalue, $\epsilon$. Previously, to solve this problem it was recommended to apply an affine transformation to the data such that the covariance matrices were identity matrices, then perform a geometric minimization [9]. Clearly, this solution is algorithmically simpler and more direct.

**Approximately Equal Covariance Matrices.** In the ideal case, the variance of each residual is unity, which poses difficulty when each covariance matrix is unique. We may, however, impose the constraint that on average, this is the case. This amounts to the assumption that $\mathsf{\Lambda}_i \approx E[\mathsf{\Lambda}]$ for $i = 1, \ldots, n$. Given the variance of the $i^{\mathrm{th}}$ residual, we compute the mean variance to be

$$E[\mathrm{Var}\,(r)] = \frac{1}{n} \sum_{i=1}^{n} \mathbf{z}_A^{\mathrm{T}} \mathsf{\Lambda}_i \mathbf{z}_A = \mathbf{z}_A^{\mathrm{T}} \left( \frac{1}{n} \sum_{i=1}^{n} \mathsf{\Lambda}_i \right) \mathbf{z}_A, \tag{27}$$

by which we define, $\mathsf{\Lambda}_{\mathrm{m}} \triangleq E[\mathsf{\Lambda}] = \frac{1}{n} \sum_{i=1}^{n} \mathsf{\Lambda}_i$. Solving the eigenvalue problem in Equation (26) with $\mathsf{\Lambda} = \mathsf{\Lambda}_{\mathrm{m}}$, will, on average, weight each error correctly.

This minimization can be justified as an approximate solution to the global minimum by the fact that it yields the exact global minimum when all covariance matrices are the same. The bias of this solution can be calculated along with the solution, since each covariance matrix deviates from the mean by $\delta\Lambda = \Lambda_{\mathrm{m}} - \Lambda_i$, and therefore the bias, $\beta_i$, of an estimate $\mathbf{z}_A$ is,

$$\beta_i = \frac{\mathbf{z}_A^{\mathrm{T}}\left(\Lambda_{\mathrm{m}} - \Lambda_i\right)\mathbf{z}_A}{\mathbf{z}_A^{\mathrm{T}}\Lambda_{\mathrm{m}}\mathbf{z}_A} = 1 - \frac{\mathbf{z}_A^{\mathrm{T}}\Lambda_i\mathbf{z}_A}{\mathbf{z}_A^{\mathrm{T}}\Lambda_{\mathrm{m}}\mathbf{z}_A}. \tag{28}$$

If $\Lambda_i \approx \Lambda_{\mathrm{m}}$, then clearly $\beta_i \approx 0$.

**Unique Covariance Matrices.** It may be that each covariance matrix varies dramatically and the bias will be large. This case requires a non-linear algorithm to find the true global minimum of the cost function $\epsilon(\mathbf{z}_A)$. The formulation of the cost function in Equation (23) enables the use of Gauss-Newton minimization.

## 4    Linearized Models

We investigate the estimation of homogeneous vectors of linearized models using the example of a circle, since it is equivalently a plane fitting problem in three dimensions. Geometrically speaking, fitting a linearized model is in general a hyperplane fitting problem. The homogeneous equation of a circle is given as

$$a\left(x^2 + y^2\right) + bx + cy + d = 0. \tag{29}$$

Substituting the model coordinates $(x, y) = (\hat{x} - \delta_x, \hat{y} - \delta_y)$ and rearranging,

$$a\left(\hat{x}^2 + \hat{y}^2\right) + b\hat{x} + c\hat{y} + d = -a\delta_x^2 - a\delta_y^2 + a\hat{x}\delta_x + a\hat{y}\delta_y + b\delta_x + c\delta_y. \tag{30}$$

That is, the functional dependence of the algebraic residual on the random variables $\delta_x$ and $\delta_y$ is described as,

$$r_{\mathrm{s}}(\delta_x, \delta_y, \hat{x}, \hat{y}, \mathbf{z}) \triangleq -a\delta_x^2 - a\delta_y^2 + a\hat{x}\delta_x + a\hat{y}\delta_y + b\delta_x + c\delta_y. \tag{31}$$

For simplicity, we assume that the errors in the point coordinates behave with covariance matrices, $\Lambda = \sigma^2 I$, although general covariance matrices may also be used. The mean value and variance of the residual error are,

$$E[r] = -2a\sigma^2 \tag{32}$$

and

$$\mathrm{Var}\left(r\right) = \mathbf{z}_A^{\mathrm{T}}C\mathbf{z}_A = \begin{bmatrix} a & b & c \end{bmatrix} \begin{bmatrix} 4\sigma^4 + \sigma^2\left(\hat{x}^2 + \hat{y}^2\right) & 2\sigma^2\hat{x} & 2\sigma^2\hat{y} \\ 2\sigma^2\hat{x} & \sigma^2 & 0 \\ 2\sigma^2\hat{y} & 0 & \sigma^2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix}. \tag{33}$$

The following are artifacts of linearizing non-linear problems, all of which undermine Gauss's least-squares theorem:

1. The variance of the residual is dependent on the measured point $(\hat{x}, \hat{y})$. That is to say, the variance of the residual depends not only on the error in the measured quantity, but also on measured quantity itself.
2. The random variables $\delta_x$ and $\delta_y$ follow Gaussian distributions, but the terms $\delta_x^2$ and $\delta_y^2$ in Equation (31) do not. Specifically they behave according to the Bessel function of the second kind, $K_0$ [10].
3. The residuals are not mean-free; this is an artifact of the non-Gaussian terms $\delta_x^2$ and $\delta_y^2$.

It is these effects which link the common direct estimation techniques to Gaussian Least Squares:

**Relation to "Normalization".** Hartley's argument for normalization was that it improves the conditioning of the design matrix [1], which is indeed true; however, this is not the true problem at hand. From a statistical point of view, normalization improves the error structure of the statistically dependent portion of the design matrix, $\mathsf{D}_A$. In consequence, the problem which normalization aims to correct (unbeknownst to its propenents) is the dependence of the variances of the residuals on the measured point. For argument's sake, say we apply normalization to the circle fitting problem. We transform the data such that the centroid is the origin, and $\frac{1}{n}\sum_{i=1}^{n}(\hat{x}_i^2 + \hat{y}_i^2) = \sqrt{2}$. If we solve the minimization with the unit norm constraint, $\mathbf{z}^{\mathrm{T}}\mathbf{z} = 1$, then the average variance of the residuals behaves according to the quadratic form, $E[\mathrm{Var}\,(r)] = \mathbf{z}^{\mathrm{T}}\mathsf{C}'\mathbf{z}$, with

$$\mathsf{C}' = \mathrm{diag}\left(4\sigma^4 + 4\sqrt{2}\sigma^2, \sigma^2, \sigma^2, 0\right) \tag{34}$$

The average variance of the residuals is consequently bounded by the eigenvalues of the matrix $\mathsf{C}'$, which are $\lambda(\mathsf{C}') = 0, \sigma^2, \sigma^2, 4\sigma^2(\sigma^2 + \sqrt{2})$. The bound on the average variance is therefore

$$0 \le E[\mathrm{Var}\,(r)] \le 4\sigma^2(\sigma^2 + \sqrt{2}). \tag{35}$$

Normalization, hence bounds the average variance of the residuals to values close to the actual noise level of the data. By this argument, circle fitting would be better implemented with data normalization followed by partitioning the statistically dependent and independent portions, which would correspond to the method of Nievergelt [14]. This would make the error behaviour of the residuals closer to an isotropic distribution; however, Equation (35) reveals the problem that the result quality would be dependent on the choice of the scaling factor.

**Relation to "Invariant Fitting".** The methods of Bookstein [2], Torr and Fitzgibbon [3], and Harker et al. [4] effectively partition quadratic terms from the linear and constant terms. That is, they effectively partition the non-Gaussian errors from the Gaussian and error free portions of the residual vectors. However, what is not treated is the fact that each residual depends on the measured point itself. This means that the residuals are each weighted irregularly (i.e. by some weighting not related to its variance), which is why the methods often lead to inappropriate fits. Effectively, the algorithm circumvents the normalization step, but is still minimizing an algebraic error.

**Relation to "Gradient Weighted Fitting".** To weight each algebraic error by the local gradient is equivalent to weighting each error by the first order Taylor approximation to its variance under the assumption of isotropic errors in the coordinates. Hence, the Gradient-Weighting scheme minimizes the algebraic error subject to the constraint that the average first order approximation to the variance is equal to unity. There is, however, an important caveat: the gradient constraint is degenerate because the derivative of the constant term is zero. This leads to gross numerical instability in the generalized eigenvectors [11], making the results heavily dependent on the conditioning and configuration of the data. This instability produces unusable results even in cases when the data exactly fits the model. Taubin himself noted that the method would yield useless results in some cases, but obviously did not correctly identify the cause as poor problem formulation [15].

## 5  Stable Direct Statistical Fitting

We previously showed that with "data normalization," the average variance of the residuals is bounded, but not constrained. In a manner analogous to fitting lines to heteroscedastic data (Section 3), we may perform the minimization subject to the constraint that the average variance is unity. This assumes that each individual covariance matrix is well approximated by the average covariance matrix. This corresponds to solving the minimization with

$$
\mathsf{D}_A = \begin{bmatrix} \hat{x}_1^2 + \hat{y}_1^2 & \hat{x}_1 & \hat{y}_1 \\ \vdots & \vdots & \vdots \\ \hat{x}_n^2 + \hat{y}_n^2 & \hat{x}_n & \hat{y}_n \end{bmatrix} \quad \text{and} \quad \mathsf{D}_B = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \triangleq \mathbf{1}, \tag{36}
$$

with the constraint matrix

$$
\mathsf{C} = \frac{1}{n} \sum_{i=1}^{n} \begin{bmatrix} 4\sigma^4 + 4\sigma^2 \left( \hat{x}_i^2 + \hat{y}_i^2 \right) & 2\sigma^2 \hat{x}_i & 2\sigma^2 \hat{y}_i \\ 2\sigma^2 \hat{x}_i & \sigma^2 & 0 \\ 2\sigma^2 \hat{y}_i & 0 & \sigma^2 \end{bmatrix}. \tag{37}
$$

The related eigenvalue problem[2] is

$$
\mathsf{D}_A^{\mathrm{T}} \left( \mathsf{I} - \mathbf{1}\mathbf{1}^+ \right) \mathsf{D}_A \mathbf{z}_A = \epsilon \mathsf{C} \mathbf{z}_A. \tag{38}
$$

This approach is numerically stable in comparison to the standard Gradient Weighting scheme by the following reasoning: For the case of a circle, some manipulation shows that, $\det \mathsf{C} = 4\sigma^6 \left( \sigma^2 + \mathrm{Var}\left( \hat{x} \right) + \mathrm{Var}\left( \hat{y} \right) \right)$. This shows that the constraint is, analytically speaking, not degenerate unless the data itself is ill-conditioned (i.e. very "point-like"). With the Gradient Weighting scheme, in contrast, the constraint is always degenerate making the eigenvectors always unstable. Worse yet, is if the data fits the model then both matrices are degenerate. This means that the algorithm is most unstable for best-case data sets, which is hardly desirable.

---

[2] This should be solved using the GSVD, but space limitations preclude just discussion.

**Summary of the Algorithm.** The algorithm can be summarized as follows, whereby steps 1 and 2 are undertaken once, offline, whereas steps 3 and 4 are the online portion of the fitting algorithm.

1. Formulate the linearized model of the geometric object, $\mathbf{d}^\mathrm{T}\mathbf{z} = 0$.
2. Partition the coefficient vector into its statistically dependent and independent portions, $\mathbf{d}_A^\mathrm{T}\mathbf{z}_A + \mathbf{d}_B^\mathrm{T}\mathbf{z}_B = 0$, such that the analytic expression for the variance of the $i^\mathrm{th}$ residual can be written as the quadratic form, $\mathrm{Var}\,(r_i) = \mathbf{z}_A^\mathrm{T}\mathsf{C}_i\mathbf{z}_A$.
3. Compute the average covariance matrix, $\mathsf{C}_\mathrm{m}$, and solve the generalized eigenvalue problem,
$$\mathsf{D}_A^\mathrm{T}\left(\mathsf{I} - \mathsf{D}_B\mathsf{D}_B^+\right)\mathsf{D}_A\mathbf{z}_A = \epsilon\mathsf{C}_\mathrm{m}\mathbf{z}_A. \tag{39}$$
4. Backsubstitute the minimizing eigenvector to find $\mathbf{z}_B = -\mathsf{D}_B^+\mathsf{D}_A\mathbf{z}_A$.

## 6   Numerical Testing

To test the new algorithm, we have applied it to the Metric Vision task of material tracking and measurement. Metric calibration of the planar scene is accomplished with circular targets to determine the homography. Figure 1 shows the scene, and the results of circle fitting. The gradient-weighted circle fit is useless due to the aforementioned numerical instability.

The position, orientation and dimensions of the steel plate can be determined by fitting a fourth order curve, or quartic. In Figure 2, the left hand images show the results of each algorithm to the edge data obtained with a contouring algorithm. The right hand images show the quartic fits after perturbing the data with a small amount of Gaussian noise ($\sigma = 1\,\mathrm{pixel}$). This test shows
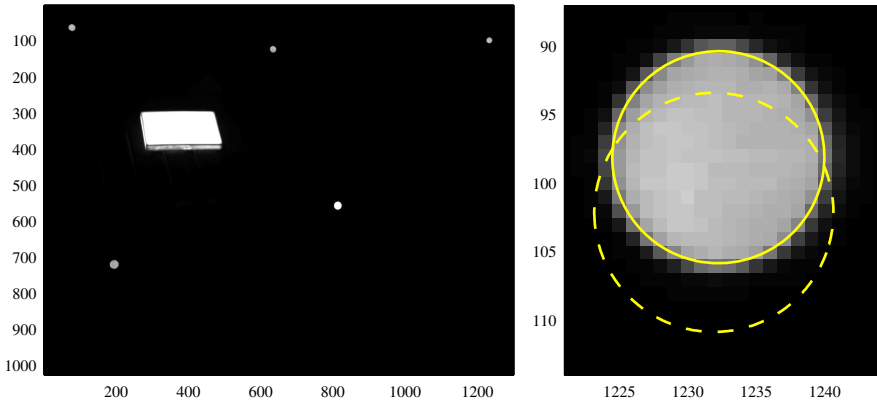


**Fig. 1.** (LEFT) The Metric Vision task of material tracking and measurement in a steel mill. (RIGHT) Circle fitting to a calibration target. The Gradient-Weighted solution $(--)$ is nonsense due to numerical instability. All other algorithms $(-)$, including a non-linear geometric fit, return the same circle.
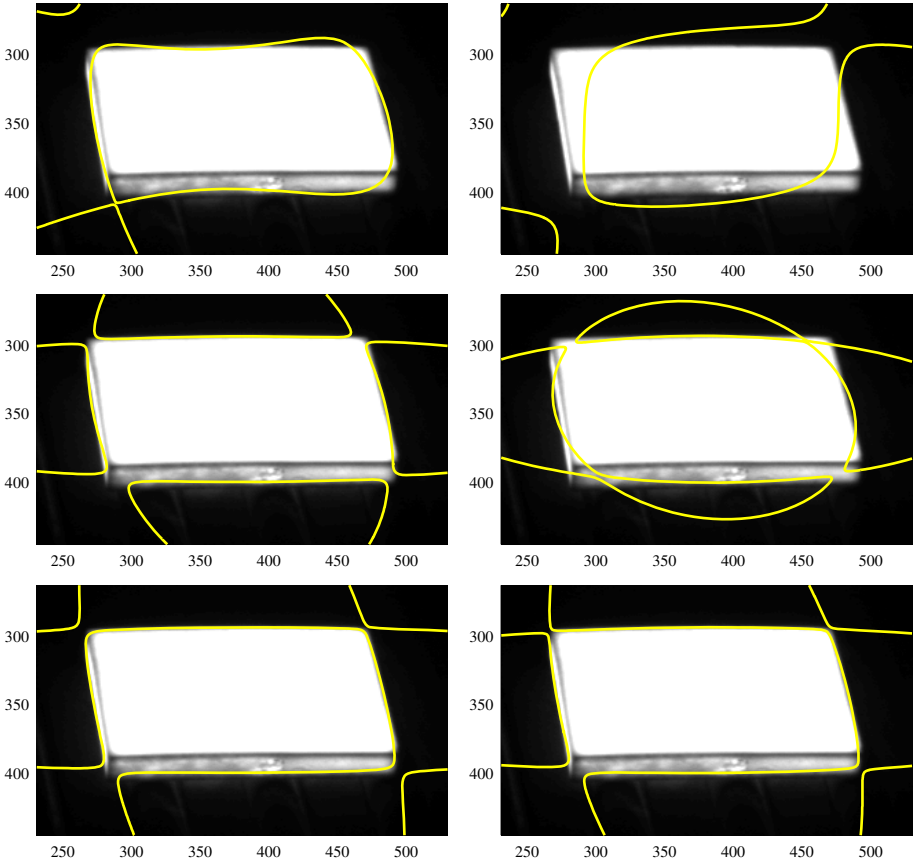
**Fig. 2.** Fitting quartic curves to determine the position, orientation, and dimensions of steel plates. Row-wise are the solutions obtained with "Gradient-Weighting", "Normalization", and the New Method. In the left column the curves are fitted to the edge data obtained by contouring. In the right column, a small amount of Gaussian noise ($\sigma = 1$ pixel) is added to demonstrate the sensitivity of each solution.

that the gradient-weighting and normalization solutions are very sensitive to perturbations in the data, which indicates instability of the solution vector. The new algorithm is not only insensitive to the large values and offsets of the image coordinates, but it is also relatively insensitive to Gaussian noise perturbing the coordinates.

## 7 Conclusion

We have proposed a generalized approach to the direct estimation of homogeneous vectors which has improved goodness-of-fit properties and numerical stability, whilst circumventing data normalization. The normal vector of the

hyperplane fit is constrained to a hyperellipsoid, which is aimed at statistically regularizing the error metric in the space of linearized models.

# References

1. Hartley, R.: In defense of the eight-point algorithm. IEEE Trans. Pattern Analysis and Machine Intelligence **19** (1997) 580–593
2. Bookstein, F.: Fitting conic sections to scattered data. Computer Graphics and Image Processing **9** (1987) 56–71
3. Torr, P., Fitzgibbon, A.: Invariant fitting of two view geometry. IEEE Trans. Pattern Analysis and Machine Intelligence **26** (2004) 648–650
4. Harker, M., O'Leary, P., Zsombor-Murray, P.: Direct type-specific conic fitting and eigenvalue bias correction. Submitted to Image and Vision Computing: 2004 British Machine Vision Conference Special Issue (2005) 17
5. Gander, W.: Least squares with a quadratic constraint. Numer. Math. **36** (1981) 291–307
6. Sampson, P.: Fitting conic sections to "very scattered" data: An iterative refinement of the Bookstein algorithm. Computer Graphics and Image Processing **18** (1982) 97–108
7. Taubin, G.: Estimation of planar curves, surfaces and nonplanar space curves defined by implicit equations with applications to edge and range image segmentation. IEEE Trans. on Pattern Analysis and Machine Intelligence **13** (1991)
8. Gauss, C.: Méthode des moindres carrés. Mémoires sur la combinaison des observations. Mallet-Bachelier, Paris (1855)
9. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Second edn. Cambridge University Press, Cambridge (2003)
10. Förstner, W.: Uncertainty and projective geometry. In Bayro Corrochano, E., ed.: Handbook of Geometric Computing. Springer (2005) 493–535
11. Golub, G., Van Loan, C.: Matrix Computations. Third edn. John Hopkins University Press, Baltimore (1996)
12. Golub, G., Hoffman, A., Stewart, G.: A generalization of the Eckart-Young-Mirsky matrix approximation theorem. Lin. Alg. and its Applic. **88/89** (1987) 317–327
13. Golub, G., Pereyra, V.: The differentiation of pseudoinverses and nonlinear least squares problems whose variables separate. SIAM J. Num. Anal. **10** (1973) 413–432
14. Nievergelt, Y.: Hyperspheres and hyperplanes fitted seamlessly by algebraic constrained total least-squares. Linear Algebra and its Applications **331** (2001) 43–59
15. Taubin, G.: An improved algorithm for algebraic curve and surface fitting. In: International Conference on Computer Vision, Berlin, Germany (1993) 658–665