Approximate Geometric Ellipsoid Fitting: A CG-Approach

Martin Kleinsteuber[1] and Knut Hüper[2]

[1] Institute of Data Processing, Technische Universität München, Germany, Kleinsteuber@tum.de
[2] Department of Mathematics, University of Würzburg, Germany Hueper@mathematik.uni-wuerzburg.de

Summary. The problem of geometric ellipsoid fitting is considered. In connection with a conjugate gradient procedure a suitable approximation for the Euclidean distance of a point to an ellipsoid is used to calculate the fitting parameters. The approach we follow here ensures optimization over the set of all ellipsoids with codimension one rather than allowing for different conics as well. The distance function is analyzed in some detail and a numerical example supports our theoretical considerations.

1 Introduction

The approximation of a set of data by an ellipsoid is an important problem in computer science and engineering, e.g. in computer vision or computer graphics, or more specifically, in 3D-reconstruction and virtual reality generation. Moreover, there are further applications in robotics [13], astronomy [18] and in metrology [2, 5, 17], as well.

Mathematically, the problem of fitting can often be expressed by a set of implicit equations depending on a set of parameters. For fixed parameters the set of equations often describes implicitly a smooth manifold, e.g. in those cases where the regular value theorem applies. The task then is to find a parameter vector, such that the corresponding manifold best fits a given set of data. As it is studied in the computer vision community, e.g. see [9, 8], a large class of computer vision problems actually falls into this category.

Certainly, there exists a variety of different ways to measure the quality of a fit, dependent on the application context. Here we focus on a certain problem of *geometric* fitting, namely, minimizing the sum of the squared Euclidean distances between the data points and the manifold. In a natural way this is a generalization of the well known linear orthogonal regression problem.

M. Diehl et al. (eds.), *Recent Advances in Optimization and its Applications in Engineering*, DOI 10.1007/978-3-642-12598-0_7, © Springer-Verlag Berlin Heidelberg 2010

A quite different approach to geometric fitting comes under the name of *algebraic* fitting which we do not follow here. It turns out that in many cases the algebraic approach has to be distinguished from the geometric one. Firstly, it seems that the numerical treatment of the former is more feasible, mainly due to the fact that the underlying optimization problem is based on a vector space model, rather than modelled in a nonlinear differential manifold setting. This might be the reason why it was preferably studied in much detail in the past, see e.g. [1, 4, 6, 10, 14, 15, 19]. Secondly, geometric fitting does not necessarily support a traditional straightforward statistical interpretation, again typical for a computer vision application, see [9] for a thorough discussion of this aspect.

For early work in the spirit of our approach, see however [11].

As already mentioned above the parameter vector might vary itself over a smooth manifold. E.g. fitting an ellipsoid of codimension one in $\mathbb{R}^n$ to a set of data points sitting in $\mathbb{R}^n$ as well, amounts in an optimization problem over the set of *all* codimension one ellipsoids. As we will see below this set can be neatly parameterized by the product of $\mathbb{R}^n$ with the set $\mathcal{P}_n$ of symmetric positive definite $n \times n$-matrices, or equivalently, by the product of $\mathbb{R}^n$ with the set $\mathcal{R}_+^{n\times n}$ of $n \times n$ upper triangular matrices with *positive* diagonal entries.

In general, there exists no explicit formula for the Euclidean distance of a point to a set. We therefore will use a suitable approximation together with a conjugate-gradient-type procedure to compute the fitting parameters.

In this paper we will put an emphasis on the geometric fitting of ellipsoids of codimension one to data points. The approach we follow here ensures that we actually optimize over all *ellipsoids* of codimension one, rather than allowing for other or even all *conics* of codimension one, or even conics of any codimension as well.

The paper is organized as follows. In the next section we motivate the quality measure we use, namely a distance function which approximates the Euclidean distance of a point to an ellipsoid in a consistent manner, in a way made precise below. We investigate the local properties of this function and compare it with the Euclidean distance and with algebraic fitting.

Differentiability of the square of this function allows for a smooth optimization procedure. In the third section we briefly describe the global parameterization of the smooth manifold of all ellipsoids of codimension one in $\mathbb{R}^n$ and set the ground for a conjugate gradient algorithm living on this manifold. The last section briefly discusses the CG-method used here, supported by a numerical example.

2 Motivation of the Distance Function

In this section we introduce a new distance measure as an approximation of the Euclidean distance from a point to an ellipsoid. This measure has the advantage that, in contrast to the Euclidean distance, it can be expressed explicitly in terms of the ellipsoid parameters and is therefore suitable for optimization tasks. Moreover, it does not have the drawback of the measure that underlies algebraic fitting, where it might happen that, given a set of points, any ellipsoid that is large enough drives the corresponding cost arbitrarily small. We specify this phenomenon in Proposition 1 below.

Let $(\cdot)^\top$ denote transposition and let

$$\mathcal{E}_{Q,\tau} := \{q \in \mathbb{R}^n \mid (q-\tau)^\top Q(q-\tau) = 1\} \tag{1}$$

be an ellipsoid with center $\tau \in \mathbb{R}^n$ and positive definite $Q \in \mathcal{P}_n$. For ellipsoids centered at the origin we shortly write $\mathcal{E}_Q := \mathcal{E}_{Q,0}$. In order to fit an ellipsoid to a given set of data $y_i \in \mathbb{R}^n$, $i = 1, \dots N$, a quality measure is required that reflects *how well* an ellipsoid fits the y_i's. There are two measures that arise in a natural way: the Euclidean distance and, since any ellipsoid defines a metric by considering it as a unit ball, the corresponding distance induced by Q. For $x, y \in \mathbb{R}^n$ denote by

$$\langle x, y \rangle_Q := x^\top Q y \tag{2}$$

the induced scalar product, the associated norm by $\|x\|_Q = (x^\top Q x)^{\frac{1}{2}}$, and the induced distance measure by

$$d_Q(x, y) := \|x - y\|_Q. \tag{3}$$

Lemma 1. *Let $x \in \mathbb{R}^n$. Then the Q-distance between x and $\mathcal{E}_Q$ is given by*

$$d_Q(x, \mathcal{E}_Q) = |1 - \|x\|_Q|. \tag{4}$$

The point of lowest Q-distance to x on $\mathcal{E}_Q$ is $\hat{x} = \frac{x}{\|x\|_Q}$.

Proof. Without loss of generality we might assume that $x \neq 0$. We compute the critical points of the function

$$a \colon \mathcal{E}_Q \to \mathbb{R}, \quad q \mapsto \|q - x\|_Q^2, \tag{5}$$

as follows. The tangent space $T_q\mathcal{E}_Q$ of $\mathcal{E}_Q$ at $q \in \mathcal{E}_Q$ is given by

$$T_q\mathcal{E}_Q := \{\xi \in \mathbb{R}^n \mid \xi^\top Q q = 0\}, \tag{6}$$

hence

$$\mathrm{D}\, a(q)\xi = 2\xi^\top Q(q - x) = -2\xi^\top Q x. \tag{7}$$

The derivative vanishes if and only if $q \in \mathbb{R}x$. A simple calculation then shows, that the minimum of a is given by

$$\hat{x} := \tfrac{x}{\|x\|_Q}. \tag{8}$$

Consequently,

$$d_Q(x, \mathcal{E}_Q) = d_Q(x, \hat{x}) = \|x - \hat{x}\|_Q = |1 - \|x\|_Q|. \tag{9}$$

□

The quality measure used in *algebraic fitting* is closely related to the Q-distance. It is defined by

$$d_{\mathrm{alg}}(x, \mathcal{E}_Q) = \left|1 - \|x\|_Q^2\right| \tag{10}$$

or, for general ellipsoids,

$$d_{\mathrm{alg}}(x, \mathcal{E}_{Q,\tau}) = \left|1 - \|x - \tau\|_Q^2\right|, \tag{11}$$

cf. [10]. Although this is easy to compute, minimizing the sum of squares of d_{alg} for a given set of noisy data points may not yield a desired result as the following proposition is stating.

Proposition 1. *Let $y_1, \dots, y_N \in \mathbb{R}^n$ be given. Then for all $\varepsilon > 0$ there exists $\delta > 0$ and $\tau \in \mathbb{R}^n$ such that*

$$\sum_{i=1}^{N} d_{\mathrm{alg}}^2(y_i, \mathcal{E}_{\delta I_n,\tau}) < \varepsilon. \tag{12}$$

Proof. Let $\delta = \delta(\tau) = \frac{1}{\|\tau\|^2}$. The claim follows since

$$\sum_{i=1}^{N} d_{\mathrm{alg}}^2(y_i, \mathcal{E}_{\delta I_n,\tau}) = \sum_{i=1}^{N}(1 - \delta\|y_i - \tau\|^2)^2 = \sum_{i=1}^{N}(1 - \tfrac{\|y_i-\tau\|^2}{\|\tau\|^2})^2 \xrightarrow{\|\tau\|\to\infty} 0.$$

□

Given a convex set $\mathcal{C} \subset \mathbb{R}^n$ and a point $x \in \mathbb{R}^n$ outside $\mathcal{C}$, it is well known that there is a unique point $q \in \partial\mathcal{C}$ on the boundary of $\mathcal{C}$ such that $d(x, \partial\mathcal{C}) = d(x, q)$, cf. Chapter 2 in [3]. If x lies in the interior of $\mathcal{C}$, this needs not to be true anymore. However, in the case where $\partial\mathcal{C} = \mathcal{E}_Q$ is an ellipsoid, q depends smoothly on x in a neighborhood of $\mathcal{E}_Q$.

Lemma 2. *Let $x \in \mathbb{R}^n$ and let $\pi\colon \mathbb{R}^n \to \mathcal{E}_Q$ be such that $d(x, \mathcal{E}_Q) = d(x, \pi(x))$. Then π is smooth in a neighborhood of $\mathcal{E}_Q$ and*

$$\mathrm{D}\,\pi(x)|_{x=q} h = \left(\mathrm{id} - \tfrac{Qqq^\top Q}{q^\top Q^2 q}\right) h. \tag{13}$$

Proof. Let $x \in \mathbb{R}^n$ be arbitrary but fixed and let $e\colon \mathcal{E}_Q \to \mathbb{R}$ with $e(q) = \frac{1}{2}\|x - q\|^2$. The minimal value of e then is $d(x, \mathcal{E}_q)$. Differentiating yields the critical point condition, namely

$$\mathrm{D}e(q)\xi = \xi^\top(q - x) = 0 \quad \text{for all} \quad \xi \in T_q\mathcal{E}_Q = \{\xi \in \mathbb{R}^n \mid \xi^\top Q q = 0\}. \tag{14}$$

Now since $T_q\mathcal{E}_Q = (\mathrm{im}(Qq))^\perp = \mathrm{im}\left(\mathrm{id} - \frac{Qqq^\top Q^\top}{q^\top Q^2 q}\right)$, the critical point condition is equivalent to

$$\left(\mathrm{id} - \tfrac{Qqq^\top Q^\top}{q^\top Q^2 q}\right)(q - x) = 0. \tag{15}$$

Using $q^\top Q q = 1$ yields

$$(q^\top Q^2 q)(q - x) - Qq + Qqq^\top Qx = 0. \tag{16}$$

Consider now the function

$$F\colon \mathcal{E}_Q \times \mathbb{R}^n \to \mathbb{R}^n, \quad (q, x) \mapsto (q^\top Q^2 q)(q - x) - Qq + Qqq^\top Qx. \tag{17}$$

Then F is smooth and $F(q, q) = 0$ for all $q \in \mathcal{E}_Q$. We use the implicit function theorem to complete the proof. The derivatives of F with respect to the first and second argument, respectively, are

$$\begin{aligned} \mathrm{D}_1 F(q,x)\xi &= (2\xi^\top Q^2 q)q + (q^\top Q^2 q)\xi - Q\xi - (2\xi^\top Q^2 q)x + Q\xi q^\top Qx + Qq\xi^\top Qx \\ \mathrm{D}_2 F(q,x)h &= Qqq^\top Qh - (q^\top Q^2 q)h. \end{aligned} \tag{18}$$

Hence $\mathrm{D}_1\, F(q, q)\xi = q^\top Q^2 q\xi$ and notice that $q^\top Q^2 q > 0$. The implicit function theorem yields the existence of a neighborhood U around q and a unique smooth function $\tilde{\pi}\colon U \to \mathcal{E}_Q$ such that $F(\tilde{\pi}(x), x) = 0$. Using π defined as above, we get $F(\pi(x), x) = 0$. Moreover, the uniqueness of $\tilde{\pi}$ implies $\tilde{\pi}|_U = \pi|_U$. Furthermore,

$$0 = \mathrm{D}\, F(\pi(x), x)h = \mathrm{D}_1\, F(\pi(x), x)\, \mathrm{D}\, \pi(x)h + \mathrm{D}_2\, F(\pi(x), x)h \tag{19}$$

and hence

$$\begin{aligned} \mathrm{D}\, \pi(x)|_{x=q} h &= -(\mathrm{D}_1\, F(\pi(q), q)^{-1}\, \mathrm{D}_2\, F(\pi(q), q)h \\ &= -(q^\top Q^2 q)^{-1}(Qqq^\top Qh - q^\top Q^2 qh) = \left(\mathrm{id} - \tfrac{Qqq^\top Q}{q^\top Q^2 q}\right) h. \end{aligned} \tag{20}$$

□

As an approximation of the Euclidean distance $d(x, \mathcal{E}_Q)$, we consider the Euclidean distance between x and $\frac{x}{\|x\|_Q}$, cf. Figure 1, i.e.

$$\tilde{d}\colon \mathbb{R}^n \setminus \{0\} \to \mathbb{R}, \quad x \mapsto \left|1 - \|x\|_Q^{-1}\right| \|x\|. \tag{21}$$

The definition of $d(x, \mathcal{E}_Q)$ immediately yields

$$d(x, \mathcal{E}_Q) \le \tilde{d}(x, \mathcal{E}_Q). \tag{22}$$

For large $\|x\|$ both d and $\tilde{d}$ tend to the same value, i.e.

$$\lim_{\|x\| \to \infty} \frac{\tilde{d}(x, \mathcal{E}_Q)}{d(x, \mathcal{E}_Q)} = 1. \tag{23}$$

An investigation of the derivatives yields the local behavior of $d, \tilde{d}$ and d_Q around some $q \in \mathcal{E}_Q$. It allows in particular to compare the first order approximations of the three distances: locally, $\tilde{d}$ behaves similar to the Euclidean

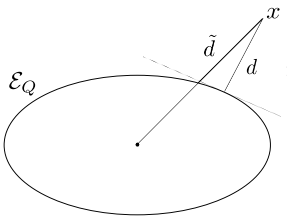

Fig. 1. Illustration of the distance measure $\tilde{d}$.

distance the more the ellipsoid becomes similar to a sphere. Moreover it shares the nice property with the Euclidean distance that it is invariant under scaling of Q, whereas the local behavior of d_Q depends on the absolute values of the eigenvalues of Q.

Proposition 2. *Let $x \in \mathbb{R}^n \setminus \{0\}$ and let $q \in \mathcal{E}_Q$. Let $\lambda_{\min}, \lambda_{\max}$ be the smallest, resp. largest eigenvalue of Q. Then*

$$\lim_{x \to q, x \notin \mathcal{E}_Q} \| \mathrm{D}\, d(x, \mathcal{E}_Q) \| = 1, \tag{24}$$

$$1 \leq \lim_{x \to q, x \notin \mathcal{E}_Q} \| \mathrm{D}\, \tilde{d}(x, \mathcal{E}_Q) \| \leq \sqrt{\tfrac{\lambda_{\max}}{\lambda_{\min}}}, \tag{25}$$

$$\sqrt{\lambda_{\min}} \leq \| \mathrm{D}\, d_Q(x, \mathcal{E}_Q) \| \leq \sqrt{\lambda_{\max}}, \quad \textit{for all } x \notin \mathcal{E}_Q, \tag{26}$$

where equality holds in the last equation either in the case of $Qx = \lambda_{\min} x$, or for $Qx = \lambda_{\max} x$.

Proof. Let $\pi(x)$ be defined as in Lemma (2), let $q \in \mathcal{E}_Q$ and let $U \subset \mathbb{R}^n$ be a neighborhood of q such that $\pi(x)$ is smooth. For $x \in U \setminus \mathcal{E}_Q$,

$$\begin{aligned} \mathrm{D}\, d(x, \mathcal{E}_Q) h &= \mathrm{D} \langle x - \pi(x), x - \pi(x) \rangle^{\frac{1}{2}} h = \left\langle h - \mathrm{D}\, \pi(x) h, \tfrac{x - \pi(x)}{\|x - \pi(x)\|} \right\rangle \\ &= \left\langle h, (\mathrm{id} - \mathrm{D}\, \pi(x))^\top \tfrac{x - \pi(x)}{\|x - \pi(x)\|} \right\rangle . \end{aligned} \tag{27}$$

Hence

$$\| \mathrm{D}\, d(x, \mathcal{E}_Q) \| = \left\| (\mathrm{id} - \mathrm{D}\, \pi(x))^\top \tfrac{x - \pi(x)}{\|x - \pi(x)\|} \right\| \leq \| (\mathrm{id} - \mathrm{D}\, \pi(x)) \|_{\mathrm{Frob}},$$

by submultiplicativity of the Frobenius norm. Therefore, using Eq. (13),

$$\lim_{x \to q, x \notin \mathcal{E}_Q} \| \mathrm{D}\, d(x, \mathcal{E}_Q) \| \leq \lim_{x \to q, x \notin \mathcal{E}_Q} \| (\mathrm{id} - \mathrm{D}\, \pi(x)) \|_{\mathrm{Frob}} = \left\| \tfrac{Qqq^\top Q}{q^\top Q^2 q} \right\|_{\mathrm{Frob}} = 1.$$

Now let

$$\gamma_x(t) = \tfrac{tx + (1-t)\pi(x)}{\|x - \pi(x)\|}.$$

Then $\pi(\gamma_x(t)) = \pi(x)$ for all $t \in (0, 1)$ and

$$d(\gamma_x(t), \mathcal{E}_Q) = d(\gamma_x(t), \pi(x)) = |t|. \tag{28}$$

Therefore, by the Cauchy-Schwarz inequality and using $\|\dot\gamma_x(t)\| = 1$,

$$\begin{aligned} 1 &= |\tfrac{\mathrm{d}}{\mathrm{d}\,t} d(\gamma_x(t), \mathcal{E}_Q)| = |\,\mathrm{D}\, d(\gamma_x(t), \mathcal{E}_Q) \cdot \dot\gamma_x(t)| \\ &\leq \|\,\mathrm{D}\, d(\gamma_x(t), \mathcal{E}_Q)\| \|\dot\gamma_x(t)\| = \|\,\mathrm{D}\, d(\gamma_x(t), \mathcal{E}_Q)\|. \end{aligned} \tag{29}$$

This proves equation (24). For Eq. (25) note that

$$\|\,\mathrm{D}\, \tilde d(x, \mathcal{E}_Q)\| = \left\| \tfrac{x}{\|x\|}(1 - \|x\|_Q^{-1}) - \|x\| \tfrac{Qx}{\|x\|_Q^3} \right\|. \tag{30}$$

The first term tends to 0 for $x \to q$ and $\|x\|_Q$ tends to 1. It is therefore sufficient to consider the term $\|x\|\|Qx\|$. Substituting $y := Q^{\frac{1}{2}}x$, which implies $\|y\|^2 \to 1$ as $x \to q$, we obtain

$$\|x\|\|Qx\| = \tfrac{(y^\top Q^{-1} y)^{\frac{1}{2}}}{\|y\|} \tfrac{(y^\top Q y)^{\frac{1}{2}}}{\|y\|} \|y\|^2 \leq \sqrt{\tfrac{\lambda_{\max}}{\lambda_{\min}}} \|y\|^2, \tag{31}$$

hence

$$\lim_{x \to q, x \notin \mathcal{E}_Q} \|\,\mathrm{D}\, \tilde d(x, \mathcal{E}_Q)\| \leq \sqrt{\tfrac{\lambda_{\max}}{\lambda_{\min}}}. \tag{32}$$

On the other hand, the Cauchy-Schwarz inequality implies

$$\lim_{x \to q} \|x\|\|Qx\| \geq \lim_{x \to q} x^\top Q x = 1. \tag{33}$$

Finally, equation (26) follows since

$$\|\,\mathrm{D}\, d_Q(x, \mathcal{E}_Q)\| = \left\| \tfrac{Qx}{\|x\|_Q} \right\| = \left(\tfrac{x^\top Q^2 x}{x^\top Q x} \right)^{\frac{1}{2}}. \tag{34}$$

□

3 Parameterization of the set of ellipsoids

Given a set of data points $y_1, \ldots y_N$, our aim is to minimize the sum of the squares of the individual distance measures $\tilde d(y_i, \mathcal{E}_{Q,\tau})$ over the set of all ellipsoids $\mathcal{E}_{Q,\tau}$, i.e. over the set

$$E := \mathcal{P}_n \times \mathbb{R}^n. \tag{35}$$

Each positive definite matrix $Q \in \mathcal{P}_n$ possesses a unique Cholesky decomposition $Q = S^\top S$, with $S \in \mathcal{R}_+^{n \times n}$, and $\mathcal{R}_+^{n\times n}$ being the set of upper triangular $n \times n$-matrices with *positive* diagonal entries. Explicit formulas for computing the Cholesky decomposition, cf. [7], imply that

$$\mathcal{R}_+^{n\times n} \to \mathcal{P}_n, \quad S \mapsto S^\top S \tag{36}$$

is a diffeomorphism. We exploit this fact to obtain a global parameterization of E. Let $\mathcal{R}^{n\times n}$ be the set of upper triangular matrices. Then $\mathcal{R}^{n\times n} \simeq \mathbb{R}^{\frac{n(n+1)}{2}}$ and

$$\phi\colon \mathcal{R}^{n\times n} \to \mathcal{R}_+^{n\times n}, \quad \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & r_{nn} \end{bmatrix} \mapsto \begin{bmatrix} \mathrm{e}^{r_{11}} & r_{12} & \cdots & r_{1n} \\ 0 & \mathrm{e}^{r_{22}} & \cdots & r_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \mathrm{e}^{r_{nn}} \end{bmatrix} \tag{37}$$

is a diffeomorphism as well. Thus

$$\mathcal{R}^{n\times n} \times \mathbb{R}^n \to E, \quad (R,\tau) \mapsto (\phi(R)^\top \phi(R), \tau) \tag{38}$$

is a *global* parameterization of the set E of codimension one ellipsoids.

4 CG-method for fitting ellipsoids to data

Using the parameterization derived in the last section and recalling that

$$\tilde{d}(x, \mathcal{E}_{Q,\tau}) = |1 - \|x-\tau\|_Q^{-1}| \cdot \|x-\tau\|,$$

a conjugate gradient method was implemented for the following problem. Given a set of data points $y_1, \ldots y_N \in \mathbb{R}^n$, minimize

$$\begin{aligned} &f\colon \mathcal{R}^{n\times n} \times \mathbb{R}^n \to \mathbb{R}, \\ &(R,\tau) \mapsto \sum_{i=1}^{N} \left(1 - \left((y_i-\tau)^\top \phi(R)^\top \phi(R)(y_i-\tau)\right)^{-\frac{1}{2}}\right)^2 \|y_i-\tau\|^2. \end{aligned} \tag{39}$$

The step-size selection was chosen using a modified one dimensional Newton step, i.e. given a point $(R,\tau) \in \mathcal{R}^{n\times n} \times \mathbb{R}^n$ and a direction $(\xi, h) \in \mathcal{R}^{n\times n} \times \mathbb{R}^n$, we have chosen the step-size

$$t^* = -\frac{\frac{\mathrm{d}}{\mathrm{d}\,t} f(R+t\xi, \tau+th)}{\left|\frac{\mathrm{d}^2}{\mathrm{d}\,t^2} f(R+t\xi, \tau+th)\right|}. \tag{40}$$

The absolute value in the denominator has the advantage, that in a neighborhood of a nondegenerated minimum the step-size coincides with the common Newton step, whereas t^* is equal to the negative of the Newton step-size if $\frac{\mathrm{d}^2}{\mathrm{d}\,t^2} f(R+t\xi, \tau+th) > 0$. Our step-size selection is also supported by simulations showing that this modification is essential for not getting stuck in local maxima or saddle points. To derive the gradient of f, for convenience we define

$$\mu_i(t) := \phi(R+t\xi)(y_i - \tau + th). \tag{41}$$

Let $\mathrm{diag}(X)$ be the diagonal matrix having the same diagonal as the matrix X and let $\mathrm{off}(X)$ be the strictly upper triangular matrix having the same upper diagonal entries as X. Then

$$\dot{\mu}_i(0) = \left(\mathrm{diag}(\xi)\,\mathrm{e}^{\mathrm{diag}(R)} + \mathrm{off}(\xi)\right)(y_i-\tau) + \phi(R)h. \tag{42}$$

Lemma 3. *Let* $\mu_i := \mu_i(0)$ *and let* $c_i := (\mu_i^\top \mu_i)^{-\frac{1}{2}}$. *The gradient of* f *evaluated at* (R, τ) *is given by*

$$\nabla f(R,\tau) = \Big(\nabla_1 f(R,\tau), \nabla_2 f(R,\tau)\Big) \tag{43}$$

where

$$\begin{aligned}\nabla_1 f(R,\tau) =& 2\sum_{i=1}^{N}(1-c_i)c_i^3\left(\mathrm{diag}((y_i-\tau)\mu_i^\top) + \mathrm{off}(\mu_i(y_i-\tau)^\top)\right),\\ \nabla_2 f(R,\tau) =& \sum_{i=1}^{N}\left(2(1-c_i)c_i^3\phi(R)^\top\mu_i + (1-c_i)^2(y_i-\tau)\right).\end{aligned} \tag{44}$$

□

The proof is lengthy but straightforward and is therefore omitted.

The algorithm was implemented using a direction update according to the formula by Polak and Ribière with restart after $n_0 := \dim E = \frac{n(n+1)}{2} + n$ steps, cf. [12]. The algorithm has the n_0*-step quadratic termination property.* That is, being a CG-method in a space diffeomorphic to a Euclidean space, it could be applied equally well to the strictly convex quadratic function $\tilde{f}(x) = x^\top Cx$ for $C \in \mathcal{P}_{n_0}$ and therefore would terminate after at most n_0 steps at the minimum of $\tilde{f}$. Consequently, under the assumption that the unique minimum of our function f is nondegenerated, the implemented CG-method is an n_0-step locally quadratic convergent algorithm, cf. [16].

In Figure 2, eight data points $y_1, \ldots, y_8$ have been generated in the following way. First, an ellipsoid $\mathcal{E}_{Q_0,\tau_0}$ has been specified and eight randomly chosen points have been normalized to $\hat{y}_1, \ldots, \hat{y}_8$, such that $\hat{y}_1, \ldots, \hat{y}_8 \in \mathcal{E}_{Q_0,\tau_0}$. Then noise has been added to obtain $y_i = \hat{y}_i + \Delta\hat{y}_i$. The figure compares the minimum of our cost function with the result of an algebraic fit (dotted line) of the y_i's. Due to Proposition 1 the algebraic fit might have a long tail.

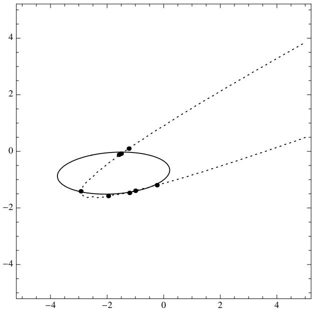

Fig. 2. Algebraic fitting (dotted line) vs. the method proposed here.

Acknowledgments

This work has been supported in parts by *CoTeSys* - Cognition for Technical Systems, Cluster of Excellence, funded by Deutsche Forschungsgemeinschaft.

References

1. F. L. Bookstein. Fitting conic sections to scattered data. *Computer Graphics and Image Processing*, 9:56–71, 1981.
2. P. Ciarlini, M. Cox, F. Pavese, and D. Richter, editors. *Advanced mathematical tools in metrology II. Proceedings of the 2nd international workshop, Oxford, UK, September 27–30, 1995.* Singapore: World Scientific, 1996.
3. F. Deutsch. *Best approximation in inner product spaces.* NY Springer, 2001.
4. A. Fitzgibbon, M. Pilu, and R. B. Fisher. Direct least square fitting of ellipses. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21(5):476–480, 1999.
5. A. Forbes. Generalised regression problems in metrology. *Numer. Algorithms*, 5(1-4):523–533, 1993.
6. W. Gander, G. Golub, and R. Strebel. Least-squares fitting of circles and ellipses. *BIT*, 34(4):558–578, 1994.
7. G. Golub and C. F. Van Loan. *Matrix computations. 3rd ed.* Baltimore: The Johns Hopkins Univ. Press, 1996.
8. K. Kanatani. *Statistical optimization for geometric computation: theory and practice.* Amsterdam: North-Holland, 1996.
9. K. Kanatani. Statistical optimization for geometric fitting: theoretical accuracy bound and high order error analysis. *Int. J. of Comp. Vision*, 80:167–188, 2008.
10. I. Markovsky, A. Kukush, and S. Van Huffel. Consistent least squares fitting of ellipsoids. *Numer. Math.*, 98(1):177–194, 2004.
11. M.Baeg, H. Hashimoto, F. Harashima, and J. Moore. Pose estimation of quadratic surface using surface fitting technique. In *Proc. of the International Conference on Intelligent Robots and Systems*, vol. 3, pages 204–209, 1995.
12. J. Nocedal and S. J. Wright. *Numerical optimization.* New York: Springer, 1999.
13. E. Rimon and S. P. Boyd. Obstacle collision detection using best ellipsoid fit. *J. Intell. Robotics Syst.*, 18(2):105–126, 1997.
14. P. D. Sampson. Fitting conic sections to 'very scattered' data: An iterative refinement of the Bookstein algorithm. *Computer Graphics and Image Processing*, 18(1):97–108, 1982.
15. S. Shklyar, A. Kukush, I. Markovsky, and S. van Huffel. On the conic section fitting problem. *J. Multivariate Anal.*, 98(3):588–624, 2007.
16. J. Stoer. On the relation between quadratic termination and convergence properties of minimization algorithms. *Numer. Math.*, 28:343–366, 1977.
17. R. Strebel, D. Sourlier, and W. Gander. A comparison of orthogonal least squares fitting in coordinate metrology. Van Huffel, Sabine (ed.), Recent advances in total least squares techniques and errors-in-variables modeling. Philadelphia: SIAM. 249-258, 1997.
18. P. Thomas, R. Binzel, M. Gaffey, B. Zellner, A. Storrs, and E. Wells. Vesta: Spin pole, size, and shape from Hubble Space Telescope images. *Icarus*, 128, 1997.
19. J. Varah. Least squares data fitting with implicit functions. *BIT*, 36(4):842–854, 1996.