Optimal Calibration of an X-Ray Detection System

KARLHEINZ SPINDLER
Fachhochschule Wiesbaden
Kurt-Schumacher-Ring 18, D-65197 Wiesbaden
GERMANY

Abstract: The calibration of an X-ray detection system for automated baggage inspection is modelled as a parameter estimation problem on a nonlinear state space. Properly taking into account the manifold structure of the state space, a calibration algorithm with extremely good convergence properties is derived.

Key Words: System identification, parameter estimation, nonlinear state space.

1 Introduction

We discuss the problem of calibrating a detection system made up of a number of individual detection devices each of which consists of a ray-emitting source and a detector made up of a row of CCDs. If a rigid body with a number of landmarks passes through such a detection device, both the times at which the individual landmarks show up in images and the resulting image coordinates can be measured. The task is to deduce from these measurements the locations of the sources and the locations and spatial orientations of the detectors. Moreover, we want to assess the accuracies of the parameter estimates obtained in terms of the (assumed) measurement accuracies and also of the body-referenced landmark coordinates if these cannot be assumed as perfectly known.

2 Detection System Model

Assume that a ray-emitting source is located at a point p and that an associated detector is centred at a point q (where the word "centred" has no specific geometric meaning; q is just a reference point or origin on the detector). Next, denote by $d := q - p$ the vector from the source to the centre of the detector and by e the unit vector specifying the direction of the detector. We extend e to a right-handed orthonormal coordinate system $(g_1 \mid g_2 \mid g_3)$ by letting

$$(1)\qquad \begin{aligned} g_1 &= e, \\ g_3 &= \frac{-d + \langle d, e\rangle e}{\|-d + \langle d, e\rangle e\|}, \\ g_2 &= g_3 \times g_1 \;=\; -\frac{d \times e}{\|d \times e\|}; \end{aligned}$$

note that $\|-d + \langle d, e\rangle e\| = \|d \times e\| = \sqrt{\|d\|^2 - \langle d, e\rangle^2}$. If a landmark located at a point a can be detected, then the ray from p through a intersects the plane through q spanned by g_1 and g_2; thus there are real numbers $\lambda > 0$, u and v such that the equation $p + \lambda(a - p) = q + ug_1 + vg_2$ holds, i.e., we have

$$(2)\qquad ug_1 + vg_2 \;=\; p - q + \lambda(a - p)$$

(where u and v are the horizontal offset and the vertical offset of the landmark image from the origin of the detection device and where necessarily $v = 0$).

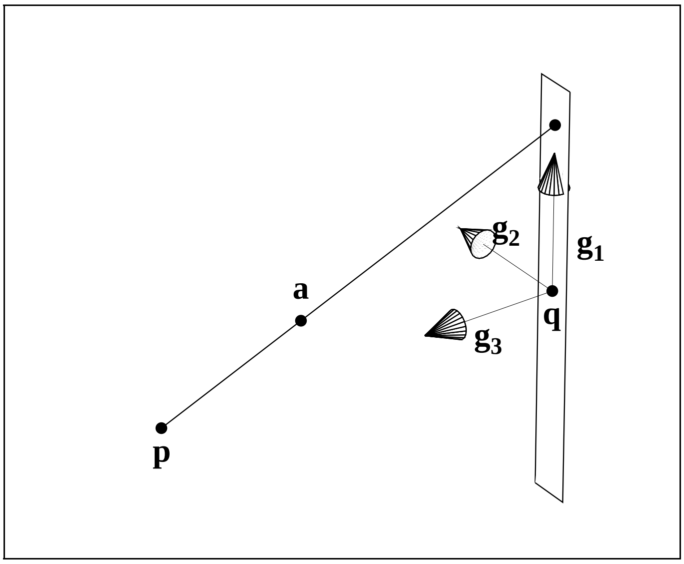

Figure 1: Detection device with linear detector.

Taking the inner product of (2) with g_3 we find that $0 = \langle p - q, g_3\rangle + \lambda\langle a - p, g_3\rangle$ and hence that

$$(3)\qquad \lambda \;=\; \frac{\langle q - p, g_3\rangle}{\langle a - p, g_3\rangle} \;=\; \frac{\langle d, g_3\rangle}{\langle a - p, g_3\rangle}\,.$$

Taking the inner product of (2) with both g_1 and g_2 and plugging in (3), we obtain the equations

$$(4)\qquad \begin{aligned} u &= \frac{\langle d, g_3\rangle\langle a - p, g_1\rangle - \langle d, g_1\rangle\langle a - p, g_3\rangle}{\langle a - p, g_3\rangle} \\ v &= \frac{\langle d, g_3\rangle\langle a - p, g_2\rangle - \langle d, g_2\rangle\langle a - p, g_3\rangle}{\langle a - p, g_3\rangle} \end{aligned}$$

which we can rewrite in the form

$$(5)\quad \begin{aligned} u &= \frac{\langle d\times(a-p), g_2\rangle}{\langle a-p, g_3\rangle} = \frac{\langle a-p, g_2\times d\rangle}{\langle a-p, g_3\rangle} \\ v &= \frac{-\langle d\times(a-p), g_1\rangle}{\langle a-p, g_3\rangle} = \frac{-\langle a-p, g_1\times d\rangle}{\langle a-p, g_3\rangle} \end{aligned}$$

using the Lagrange identity $\langle a\times b, c\times d\rangle = \langle a,c\rangle\langle b,d\rangle - \langle a,d\rangle\langle b,c\rangle$. The condition $v=0$ implies

$$(6)\quad \langle a-p, g_1\times d\rangle = 0.$$

In the next section the measurement equations will be derived for the special situation that a rigid body with identifiable landmarks moves along a conveyor belt and passes through the detection system considered. In this case we fix a space-fixed reference coordinate system (e_1, e_2, e_3) in which e_3 is the direction of the motion.

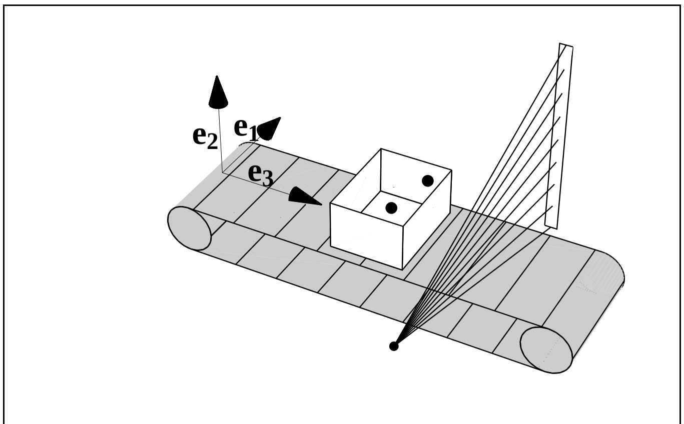

Figure 2: Reference system and moving landmarks.

3 Measurement Equations

Assume that a rigid body (such as a suitcase) moves along the conveyor belt with constant speed s. We fix a body-fixed reference point and a body-fixed coordinate system (a, b) on the bottom side of the body; moreover, we denote by φ the angle which a makes with e_3, so that

$$(7)\quad \begin{aligned} a &= \cos\varphi\, e_3 + \sin\varphi\, e_1, \\ b &= -\sin\varphi\, e_3 + \cos\varphi\, e_1. \end{aligned}$$

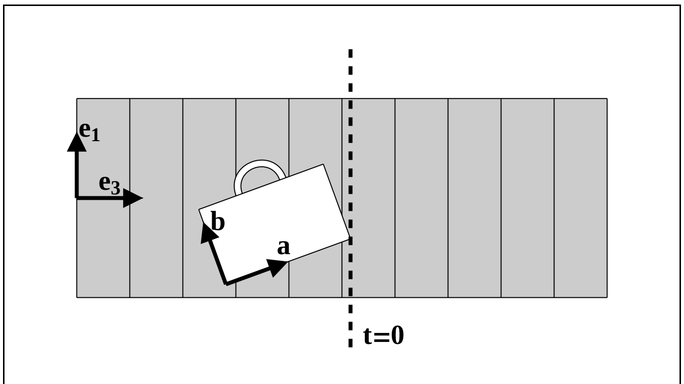

Figure 3: View from top.

Next, we fix a reference time and denote by $a_{\text{ref}} = x_{\text{ref}} e_1 + z_{\text{ref}} e_3$ the space-referenced coordinates of the body reference point at this time. We now consider a landmark whose body coordinates are given by the distances A and B and its height H; then the spatial position of this landmark at the reference time is given by $a_0 = a_{\text{ref}} + A\cdot a + B\cdot b + H\cdot e_2$, i.e., by

$$(8)\quad a_0 = \begin{bmatrix} x_{\text{ref}} + A\sin\varphi + B\cos\varphi \\ H \\ z_{\text{ref}} + A\cos\varphi - B\sin\varphi \end{bmatrix}.$$

Consequently, the landmark position at the time of detection is $a = a_0 + tse_3$ where s is the speed with which the conveyor belt moves and where, due to (6), the detection time t is determined by the condition

$$(9)\quad 0 = \langle a_0 - p, g_1\times d\rangle + ts\langle e_3, g_1\times d\rangle$$

and hence (writing $e = g_1$) is given by

$$(10)\quad \begin{aligned} t &= -\frac{\langle a_0-p, e\times d\rangle}{s\langle e_3, e\times d\rangle} = -\frac{\langle (a_0-p)\times d, e\rangle}{s\langle e_3\times d, e\rangle} \\ &= -\frac{\langle (a_0-p)\times e, d\rangle}{s\langle e_3\times e, d\rangle} =: T(p,d,e,a_0,s). \end{aligned}$$

Thus

$$(11)\quad \begin{aligned} a-p &= a_0 - p + tse_3 \\ &= a_0 - p - \frac{\langle a_0-p, e\times d\rangle}{\langle e_3, e\times d\rangle} e_3 \\ &= \frac{\langle e_3, e\times d\rangle(a_0-p) - \langle a_0-p, e\times d\rangle e_3}{\langle e_3, e\times d\rangle} \\ &= \frac{(e\times d)\times\big((a_0-p)\times e_3\big)}{\langle e_3, e\times d\rangle} \\ &= \frac{\langle (a_0-p)\times e_3, e\rangle d - \langle (a_0-p)\times e_3, d\rangle e}{\langle e_3, e\times d\rangle} \\ &= \frac{\langle w,e\rangle d - \langle w,d\rangle e}{\langle e_3, e\times d\rangle} \end{aligned}$$

where

$$(12)\quad w := (a_0-p)\times e_3\,.$$

Plugging (1) into (5) and using (11), we obtain the measurement value

$$(13)\quad \begin{aligned} u &= \frac{\big\langle a-p,\; d\times(d\times e)\big\rangle}{\big\langle a-p,\; -d+\langle d,e\rangle e\big\rangle} \\ &= \frac{\big\langle \langle w,e\rangle d - \langle w,d\rangle e, \langle d,e\rangle d - \|d\|^2 e\big\rangle}{\big\langle \langle w,e\rangle d - \langle w,d\rangle e, -d + \langle d,e\rangle e\big\rangle} \\ &= \frac{\langle w,d\rangle\big(\|d\|^2 - \langle d,e\rangle^2\big)}{\langle w,e\rangle\big(\langle d,e\rangle^2 - \|d\|^2\big)} \\ &= -\frac{\langle w,d\rangle}{\langle w,e\rangle} = -\frac{\langle (a_0-p)\times e_3, d\rangle}{\langle (a_0-p)\times e_3, e\rangle} \\ &=: U(p,d,e,a_0)\,. \end{aligned}$$

Thus the functions T and U express the time and value of the measurement in terms of the parameters p, d, e, s and a_0 where, in turn, a_0 is a function of the body coordinates A, B, H and the displacement coordinates $x_{\text{ref}}, y_{\text{ref}}, \varphi$.

4 Partial Derivatives

To see how sensitively the functions T und U depend on their arguments, we calculate the associated partial derivatives. We first determine the gradients of T and U with respect to p, d and e (where e is considered as an element of $\mathbb{R}^3$, i.e., where the constraint $\|e\| = 1$ is momentarily ignored). Starting with T, we find first that

$$\nabla_p T = \frac{d \times e}{s\langle e_3 \times d, e\rangle}. \tag{14}$$

Next we see that $\nabla_d T$ equals

$$\begin{aligned} &\frac{\langle (a_0-p) \times e, d\rangle (e_3 \times e) - \langle e_3 \times e, d\rangle \big((a_0-p) \times e\big)}{s\langle e_3 \times e, d\rangle^2} \\ &= \frac{\Big(\langle a_0-p, e \times d\rangle e_3 - \langle e_3, e \times d\rangle (a_0-p)\Big) \times e}{s\langle e_3 \times e, d\rangle^2} \\ &= \frac{\Big((e \times d) \times \big(e_3 \times (a_0-p)\big)\Big) \times e}{s\langle e_3 \times e, d\rangle^2} \\ &= \frac{\big(w \times (e \times d)\big) \times e}{s\langle e_3 \times e, d\rangle^2} = \frac{\langle w, e\rangle (e \times d)}{s\langle e_3 \times e, d\rangle^2} \end{aligned} \tag{15}$$

so that

$$\nabla_d T = \frac{\langle w, e\rangle (e \times d)}{s\langle e_3 \times e, d\rangle^2} = \frac{\langle w, e\rangle (e \times d)}{s\langle e_3 \times d, e\rangle^2}. \tag{16}$$

Finally, $\nabla_e T$ equals

$$\begin{aligned} &\frac{\langle (a_0-p) \times d, e\rangle (e_3 \times d) - \langle e_3 \times d, e\rangle \big((a_0-p) \times d\big)}{s\langle e_3 \times d, e\rangle^2} \\ &= \frac{\Big(\langle a_0-p, d \times e\rangle e_3 - \langle e_3, d \times e\rangle (a_0-p)\Big) \times d}{s\langle e_3 \times d, e\rangle^2} \\ &= \frac{\Big((d \times e) \times \big(e_3 \times (a_0-p)\big)\Big) \times d}{s\langle e_3 \times d, e\rangle^2} \\ &= \frac{\big(w \times (d \times e)\big) \times d}{s\langle e_3 \times d, e\rangle^2} = \frac{\langle w, d\rangle (d \times e)}{s\langle e_3 \times d, e\rangle^2} \end{aligned} \tag{17}$$

so that

$$\nabla_e T = \frac{\langle w, d\rangle (d \times e)}{s\langle e_3 \times d, e\rangle^2}. \tag{18}$$

Turning to U, we find that

$$\begin{aligned} \nabla_p U &= \frac{\langle w, e\rangle (e_3 \times d) - \langle w, d\rangle (e_3 \times e)}{\langle w, e\rangle^2} \\ &= \frac{e_3 \times \big(\langle w, e\rangle d - \langle w, d\rangle e\big)}{\langle w, e\rangle^2} \\ &= \frac{e_3 \times \big(w \times (d \times e)\big)}{\langle w, e\rangle^2} \end{aligned} \tag{19}$$

and also

$$\nabla_d U = \frac{-w}{\langle w, e\rangle} \quad \text{and} \quad \nabla_e U = \frac{\langle w, d\rangle w}{\langle w, e\rangle^2}. \tag{20}$$

Moreover, we obviously have

$$\nabla_{a_0} T = -\nabla_p T \quad \text{and} \quad \nabla_{a_0} U = -\nabla_p U; \tag{21}$$

the partial derivatives of T and U with respect to any of the parameters A, B, H and $x_{\text{ref}}, y_{\text{ref}}, \varphi$ are then given by

$$\frac{\partial T}{\partial \bullet} = \langle \nabla_{a_0} T, \frac{\partial a_0}{\partial \bullet}\rangle \quad \text{and} \quad \frac{\partial U}{\partial \bullet} = \langle \nabla_{a_0} U, \frac{\partial a_0}{\partial \bullet}\rangle \tag{22}$$

where

$$\begin{aligned} &\frac{\partial a_0}{\partial A} = \begin{bmatrix} \sin\varphi \\ 0 \\ \cos\varphi \end{bmatrix}, \quad \frac{\partial a_0}{\partial B} = \begin{bmatrix} \cos\varphi \\ 0 \\ -\sin\varphi \end{bmatrix}, \\ &\frac{\partial a_0}{\partial H} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \frac{\partial a_0}{\partial \varphi} = \begin{bmatrix} A\cos\varphi - B\sin\varphi \\ 0 \\ -A\sin\varphi - B\cos\varphi \end{bmatrix}, \\ &\frac{\partial a_0}{\partial x_{\text{ref}}} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \frac{\partial a_0}{\partial z_{\text{ref}}} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \end{aligned} \tag{23}$$

Finally,

$$\frac{\partial T}{\partial s} = \frac{\langle (a_0-p) \times d, e\rangle}{s^2\langle e_3 \times d, e\rangle} = \frac{-T}{s}. \tag{24}$$

We now use the partial derivatives to show which changes in the measurement functions T and U are caused by changes in the system parameters. The only parameter for which this is not straightforward is the direction e, because now we have to incorporate the constraint $\|e\| = 1$. (In other words, the argument e of the functions T and U must not be considered as an element of the linear manifold $\mathbb{R}^3$, but as an element of the nonlinear manifold $\mathbb{S} := \{x \in \mathbb{R}^3 \mid \|x\| = 1\}$.) One possible way of proceeding would be to choose a parametrisation

$$(\theta, \varphi) \mapsto \begin{bmatrix} \xi(\theta, \varphi) \\ \eta(\theta, \varphi) \\ \zeta(\theta, \varphi) \end{bmatrix} \tag{25}$$

of the unit sphere $\mathbb{S}$, for example spherical polar coordinates

$$\begin{bmatrix} \xi(\theta,\varphi) \\ \eta(\theta,\varphi) \\ \zeta(\theta,\varphi) \end{bmatrix} = \cos\theta\cos\varphi\, b_1 + \cos\theta\sin\varphi\, b_2 + \sin\theta\, b_3 \tag{26}$$

with respect to a reference system (b_1, b_2, b_3) chosen such that the parametrisation is smoothly invertible about the actual unit vector e, and then to calculate the partial derivatives with respect to the coordinates θ and ϕ. Proceeding this way, however, may lead to numerical difficulties; therefore, we choose a different approach. Namely, we consider only increments $\delta e \in \mathbb{R}^3$ which are tangent to $\mathbb{S}$ at e (with the consequence that $e + \delta e$ is an element of the unit sphere $\mathbb{S}$ up to second-order effects). Let us write

$$e = \begin{bmatrix} \xi \\ \eta \\ \zeta \end{bmatrix} \tag{27}$$

where $\xi^2 + \eta^2 + \zeta^2 = 1$; since the detector cannot be aligned with the direction of motion of the conveyor belt, we have $e \neq e_3$ and hence $\xi^2 + \eta^2 \neq 0$. Then the tangent space $T_e\mathbb{S}$ of $\mathbb{S}$ at the point e is spanned by the vectors

$$\begin{aligned} T_1(e) &:= \frac{1}{\sqrt{\xi^2+\eta^2}} \begin{bmatrix} -\eta \\ \xi \\ 0 \end{bmatrix} \quad \text{and} \\ T_2(e) &:= e \times T_1(e) = \frac{1}{\sqrt{\xi^2+\eta^2}} \begin{bmatrix} -\xi\zeta \\ -\eta\zeta \\ \xi^2+\eta^2 \end{bmatrix} \end{aligned} \tag{28}$$

so that we must consider only increments of the form

$$\delta e = \Delta_1 \cdot T_1(e) + \Delta_2 \cdot T_2(e). \tag{29}$$

Note that the fact that $T_1(e)$ and $T_2(e)$ are orthonormal unit vectors implies that

$$\|\delta e\| = \sqrt{\Delta_1^2 + \Delta_2^2}\,. \tag{30}$$

Now changing e by an increment δe as in (29) yields, in first-order approximation, changes $(\delta T)_e$ in T and $(\delta U)_e$ in U which, using (18) and (20) and the identities

$$\begin{aligned} \langle d\times e, T_1(e)\rangle &= \langle d, e\times T_1(e)\rangle = \langle d, T_2(e)\rangle, \\ \langle d\times e, T_2(e)\rangle &= \langle d\times e, e\times T_1(e)\rangle = -\langle d, T_1(e)\rangle, \end{aligned} \tag{31}$$

are given by

$$\begin{aligned} (\delta T)_e &= \langle \nabla_e T, \delta e\rangle = \langle \nabla_e T, \Delta_1 T_1(e) + \Delta_2 T_2(e)\rangle \\ &= \frac{\langle w,d\rangle\langle d,T_2(e)\rangle}{s\langle e_3\times d, e\rangle^2}\Delta_1 - \frac{\langle w,d\rangle\langle d,T_1(e)\rangle}{s\langle e_3\times d, e\rangle^2}\Delta_2 \end{aligned} \tag{32}$$

and

$$\begin{aligned} (\delta U)_e &= \langle \nabla_e U, \delta e\rangle = \langle \nabla_e U, \Delta_1 T_1(e) + \Delta_2 T_2(e)\rangle \\ &= \frac{\langle w,d\rangle\langle w,T_1(e)\rangle}{\langle w,e\rangle^2}\Delta_1 + \frac{\langle w,d\rangle\langle w,T_2(e)\rangle}{\langle w,e\rangle^2}\Delta_2. \end{aligned} \tag{33}$$

5 Estimation Procedure

The practical determination of the detector configuration from the available measurements requires filtering out the noise with which the measurements are fraught. This is a standard estimation problem which, in reasonable generality, can be formulated as follows. (See [2], pp. 120-133 for more details.) A measurement vector μ depends on two kinds of parameters U and u which are distinguished because of the different roles they play in the subsequent estimation process: U is treated as a *solve-for parameter* whereas u is taken as a *consider parameter*; i.e., the value of U will be estimated whereas u is only considered in assessing the accuracy of the estimate obtained for U. (In our case $U = (p, d, e, \varphi, s)$ while $u = (A_1, B_1, H_1, \ldots, A_N, B_N, H_N)$ where (A_i, B_i, H_i) are the body-referenced coordinates of the i-th landmark, where $1 \leq i \leq N$.) Alternatively, φ and s can also be treated as consider rather than solve-for parameters.) If $U^\star$ and $u^\star$ are the true (but unknown) parameter values then the measurement vector $\widehat{\mu}$ obtained is

$$\widehat{\mu} = \mu(U^\star, u^\star) + n \tag{34}$$

where n is the measurement noise (whose covariance matrix is supposed to be known). We assume that we have initial estimates U_{init} and u_{init} for the parameters in question. While the estimate for u is never changed, we want to iteratively improve the available estimate for U. Thus we ask how to optimally update an "old" estimate U_{old} to obtain a "new" estimate

$$U_{\text{new}} = U_{\text{old}} + \delta U\,. \tag{35}$$

To assess the quality of an arbitrary estimate (U, u), we introduce the *residual vector*

$$\rho(U,u) = \widehat{\mu} - \mu(U,u) \tag{36}$$

which is a list of the differences between the actually obtained and the theoretically expected measurements. To properly measure the size of the residual vector, we weight the different measurements according to their respective accuracies; i.e., we introduce the scalar quantity

$$Q(U,u) := \rho(U,u)^T W \rho(U,u) \tag{37}$$

with the weighting matrix

$$W = \mathrm{Cov}[n]^{-1}\,. \tag{38}$$

Denoting by $\sqrt{W}$ the unique upper triangular matrix M such that $W = M^T M$ (obtained by performing the *Cholesky decomposition* of W; see [1], pp. 37-43, and [3], pp. 146-149) we can write

$$Q(U,u) \;=\; \| \sqrt{W}\rho(U,u) \|^2; \tag{39}$$

thus in the case of uncorrelated measurements Q is simply the sum of the squares of the weighted residuals, where the weighting factor for any measurement is the reciprocal of the standard deviation of this measurement. Now an update step δU as in (35) is considered optimal if it minimises the size of the resulting "new" residual vector

$$\begin{aligned} \rho_{\text{new}} &= \rho(U_{\text{new}}, u_{\text{init}}) \;=\; \rho(U_{\text{old}} + \delta U, u_{\text{init}}) \\ &\approx \rho(U_{\text{old}}, u_{\text{init}}) + (\partial\rho/\partial U)(U_{\text{old}}, u_{\text{init}})\delta U \\ &= \rho_{\text{old}} - A(U_{\text{old}}, u_{\text{init}})\delta U \end{aligned} \tag{40}$$

where

$$A(U,u) \;:=\; \frac{\partial\mu}{\partial U}(U,u) \tag{41}$$

denotes the matrix of partial derivatives of the measurements with respect to the solve-for parameters. Thus, using first-order approximations, we want to choose the update δU such that

$$Q_{\text{new}} = \| \sqrt{W}\rho_{\text{new}} \|^2 = \| \sqrt{W}\rho_{\text{old}} - \sqrt{W}A\delta U \|^2 \tag{42}$$

(where $A := A(U_{\text{old}}, u_{\text{init}})$) becomes minimal. It is well known (see [2], pp. 109-119) that if A has maximal rank this minimisation problem has the unique solution

$$\delta U \;=\; (A^T W A)^{-1} A^T W \rho_{\text{old}}\,. \tag{43}$$

However, the matrix $A^T W A$ is often ill-conditioned; thus for numerical reasons it is not recommended to perform the matrix inversion in (43) in a straightforward way. Instead, we determine an orthogonal matrix P such that

$$P\sqrt{W}A \;=:\; R \;=\; \begin{bmatrix} R_1 \\ 0 \end{bmatrix} \tag{44}$$

has upper triangular form (where R_1 is an upper triangular square matrix whose size is given by the number of solve-for parameters). (Such a matrix P can be determined by a sequence of *Householder transformations*; see [1], pp. 57-67, and [3], pp. 164-168.) We let

$$\xi := P\sqrt{W}\rho_{\text{old}}; \tag{45}$$

since applying an orthogonal matrix does not effect the norm of a vector, (42) becomes

$$\begin{aligned} Q_{\text{new}} &= \|\xi - R\,\delta U\|^2 \;=\; \left\| \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} - \begin{bmatrix} R_1\delta U \\ 0 \end{bmatrix} \right\|^2 \\ &= \|\xi_1 - R_1\delta U\|^2 + \|\xi_2\|^2; \end{aligned} \tag{46}$$

it is clear that this last expression is minised by letting

$$\delta U \;:=\; R_1^{-1}\xi_1. \tag{47}$$

Note that (47) yields (43) because

$$\begin{aligned} &R_1^{-1}\xi_1 \;=\; (R_1^T R_1)^{-1} R_1^T \xi_1 \;=\; (R^T R)^{-1} R^T \xi \;= \\ &(A^T \sqrt{W}^T P^T P \sqrt{W} A)^{-1} A^T \sqrt{W}^T P^T P \sqrt{W} \rho_{\text{old}} \end{aligned} \tag{48}$$

which, using $P^T P = \mathbf{1}$, becomes $(A^T W A)^{-1} A^T W \rho_{\text{old}}$. Thus we know how the update step (35) should be performed. Since in each step we linearised about the current estimate, iteration of the procedure is necessary. To monitor convergence, we note from (40) that we can predict which residual vector can be expected in the next iteration (to be performed with U_{new} instead of U_{old}), namely

$$\rho_{\text{expected}} \;=\; \rho_{\text{old}} - A(U_{\text{old}}, u_{\text{init}})\delta U\,. \tag{49}$$

We consider convergence to be achieved if the difference between the residual vectors expected for and actually obtained in the next iteration becomes "small"; i.e., if

$$\max_{1\le i\le N} \left| \left(\sqrt{W}(\rho_{\text{obtained}} - \rho_{\text{expected}}) \right)_i \right| \;<\; \varepsilon \tag{50}$$

for some predefined convergence margin $\varepsilon > 0$. It remains to assess the accuracy of the estimate obtained. After convergence, all remaining residuals are supposed to stem exclusively from the measurement noise and the uncertainty in the consider parameter estimate u_{init} (whereas the final estimate obtained for U is supposed to be the true value $U^\star$). Then, if $\delta u := u_{\text{init}} - u^\star$ is the error in the estimate for u, the residual vector becomes

$$\begin{aligned} \rho &= \widehat{\mu} - \mu(U^\star, u^\star + \delta u) \\ &= \mu(U^\star, u^\star) + n - \mu(U^\star, u^\star + \delta u) \\ &\approx \mu(U^\star, u^\star) + n - \mu(U^\star, u^\star) - (\partial\mu/\partial u)(U^\star, u^\star)\delta u \\ &= n - (\partial\mu/\partial u)(U^\star, u^\star)\delta u \\ &\approx n - (\partial\mu/\partial u)(U^\star, u_{\text{init}})\delta u\,. \end{aligned} \tag{51}$$

Making the (natural) assumption that the measurement noise and the error in the consider parameter estimate are uncorrelated and writing

$$B \;:=\; \frac{\partial\mu}{\partial u}(U^\star, u_{\text{init}}), \tag{52}$$

we find from (51) that

$$\begin{aligned} \text{Cov}[\rho] &= \text{Cov}[n] + \text{Cov}[B\,\delta u] \\ &= W^{-1} + B\,\text{Cov}[\delta u]B^T \end{aligned} \tag{53}$$

and hence from (43) that

$$\begin{aligned} \mathrm{Cov}[\delta U] &= (A^T WA)^{-1} A^T W \,\mathrm{Cov}[\rho] WA (A^T WA)^{-1} \\ &= (A^T WA)^{-1} + D\,\mathrm{Cov}[\delta u] D^T \end{aligned} \tag{54}$$

where $D := (A^T WA)^{-1}(A^T WB)$. Note that the first summand in (54) represents the parameter estimation inaccuracy due to the noise in the measurements whereas the second summand represents the parameter estimation inaccuracy due the consider parameter uncertainty.

6 Nonlinear Update

When the general estimation procedure explained in the previous section is applied to the calibration problem at hand, a peculiarity arises. Namely, the parameter e (and hence the solve-for parameter $U = (p, d, e, \varphi, s)$) cannot be taken as an element of a linear manifold, due to the constraint $\|e\| = 1$. As stated at the end of section 4, this is dealt with by allowing only updates of the special form (29). This has the consequence that a typical row of the partial derivative matrix (41) takes the form

$$\left((\nabla_p m)^T, (\nabla_d m)^T, \langle \nabla_e m, T_1 \rangle, \langle \nabla_e m, T_2 \rangle, \frac{\partial m}{\partial \varphi}, \frac{\partial m}{\partial s} \right) \tag{55}$$

where m is the measurement in question; the update vector then takes the form

$$(\delta p_1,\, \delta p_2,\, \delta p_3,\, \delta d_1,\, \delta d_2,\, \delta d_3,\, \Delta_1,\, \Delta_2,\, \delta\varphi,\, \delta s)^T\,. \tag{56}$$

Once the update vector is found, we can form δe via (29), but, of course, we cannot simply write $e_{\text{new}} = e_{\text{old}} + \delta e$ because the right-hand side of this equation is not an element of $\mathbb{S}$. What we do instead is "wrap around" the vector δe and hence apply the update along the geodesic of $\mathbb{S}$ originating from e_{old} determined by the tangent vector δe.

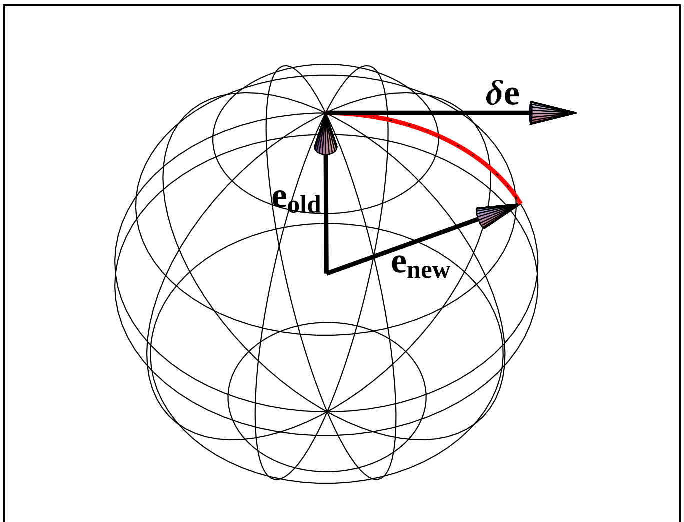

Figure 4: Nonlinear update of direction vector.

This results in the nonlinear update step

$$e_{\text{new}} \;:=\; \cos(\|\delta e\|)\, e_{\text{old}} + \frac{\sin(\|\delta e\|)}{\|\delta e\|}\delta e. \tag{57}$$

After convergence, the computed covariance matrix involving (Δ_1, Δ_2) can be easily converted into the covariance matrix involving the unit vector $e = (\xi, \eta, \zeta)^T$ by introducing the (3×2)-matrix

$$\widehat{T} \;:=\; \begin{bmatrix} | & | \\ T_1(e) & T_2(e) \\ | & | \end{bmatrix} \tag{58}$$

whose two columns are just the tangent vectors $T_i(e)$. Writing $\Delta := (\Delta_1, \Delta_2)^T$, equation (29) takes the form

$$\delta e \;=\; \widehat{T}\,\Delta \tag{59}$$

which implies

$$\mathrm{Cov}[\delta e] \;=\; \widehat{T}\,\mathrm{Cov}[\Delta]\,\widehat{T}^T\,. \tag{60}$$

Conversely we have $\widehat{T}^T \mathrm{Cov}[\delta e]\widehat{T} = \widehat{T}^T\widehat{T}\,\mathrm{Cov}[\Delta]\widehat{T}\widehat{T}^T$ which, using $\widehat{T}^T\widehat{T} = \mathbf{1}$, is just $\mathrm{Cov}[\Delta]$; hence

$$\mathrm{Cov}[\Delta] \;=\; \widehat{T}^T\,\mathrm{Cov}[\delta e]\,\widehat{T}\,. \tag{61}$$

Thus if C is the (10×10) covariance matrix for the vector (56), the (11×11) covariance matrix for the estimate (p, d, e, φ, s) is given by $\Theta C \Theta^T$ where Θ is the (11×10) matrix

$$\Theta \;:=\; \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \widehat{T}_{11} & \widehat{T}_{12} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \widehat{T}_{21} & \widehat{T}_{22} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \widehat{T}_{31} & \widehat{T}_{32} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}; \tag{62}$$

i.e.,

$$\Theta \;:=\; \begin{bmatrix} \mathbf{1}_6 & 0 & 0 \\ 0 & \widehat{T} & 0 \\ 0 & 0 & \mathbf{1}_2 \end{bmatrix} \tag{63}$$

where, in general, we denote by $\mathbf{1}_m$ the $(m \times m)$ identity matrix.

7 Examples

As examples we present two test cases in which 30 landmarks were used whose locations within a suitcase of length 800 mm and of width 600 mm are shown in the following figure. (Each black circle represents two landmarks at heights of 10 mm and 200 mm above the conveyor belt whereas each white circle represents two landmarks at heights of 50 mm and 120 mm.)

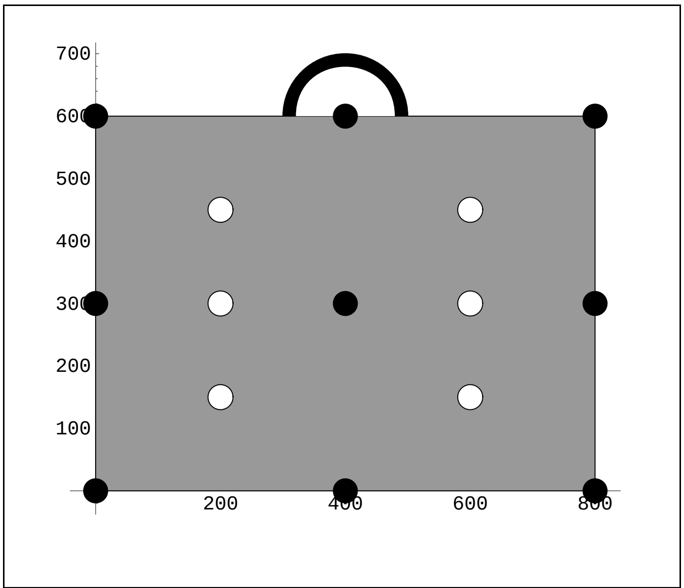

Figure 5: Landmarks used in test cases.

We assumed measurement accuracies of 1.5 mm for each u-coordinate (reflecting an image processing accuracy of one pixel) and a timing accuracy of $350^{-1}s$ (being the reciprocal of the scanning frequency of 350 Hz). We made the rather pessimistic assumption that each of the position coordinates of each of the landmarks is fraught with noise which is normally distributed with a standard deviation of 5 mm. Initial errors of 30 mm for each position coordinate and of 5 degrees for each angular variable were introduced; morever, the correct value of 500 mm/s for the speed of the conveyor belt was increased by 10%. In the first example we used one single detector pointing in a direction perpendicular to the conveyor belt; see Figure 2. The estimates obtained in the various iterations were as follows.

parameter	true	0	1
p_1 (mm)	-700.0	-730.0	-699.7
p_2 (mm)	200.0	170.0	207.2
p_3 (mm)	1000.0	970.0	992.0
d_1 (mm)	2100.0	2070.0	2053.8
d_2 (mm)	0.0	-30.0	-11.8
d_3 (mm)	0.0	-30.0	84.7
ξ	0.0	0.0354	-0.131
η	1.0	0.9929	0.9913
ζ	0.0	0.1134	-0.008
φ (deg)	20.0	25.0	17.3
s (mm/s)	500.0	550.0	495.8
RMS		1228.4	864.1
RESDIF			786.0

parameter	2	3	4
p_1 (mm)	-700.5	-700.7	-700.8
p_2 (mm)	200.2	199.9	199.9
p_3 (mm)	995.4	999.5	999.7
d_1 (mm)	2066.3	2096.2	2097.8
d_2 (mm)	1.8	1.0	0.9
d_3 (mm)	20.5	0.6	-0.4
ξ	-0.050	-0.008	-0.005
η	0.9987	1.0000	1.0000
ζ	0.0014	-0.000	-0.001
φ (deg)	19.4	20.0	20.0
s (mm/s)	499.3	499.5	499.5
RMS	174.4	16.6	7.1
RESDIF	165.4	14.4	0.1

In a second example we used an L-shaped detector consisting of two rows of CCDs, one pointing perpendicularly to the conveyor belt, the other being placed transversally across the conveyor belt; see Figure 6.

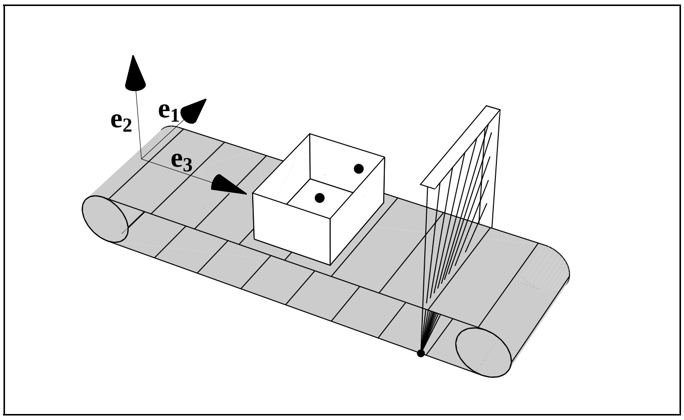

Figure 6: Detector geometry in the second example.

In this example the following estimation results were obtained.

parameter	true	0	1
p_1 (mm)	-500.0	-530.0	-497.4
p_2 (mm)	-200.0	-230.0	-217.5
p_3 (mm)	1000.0	970.0	999.0
d_1 (mm)	1000.0	970.0	991.5
d_2 (mm)	1200.0	1170.0	1267.5
d_3 (mm)	0.0	-30.0	50.5
ξ_1	0.0	-.1077	.0090
η_1	-1.0	-.9932	-.9995
ζ_1	0.0	.0451	-.0288
ξ_2	-1.0	-.9929	-.9991
η_2	0.0	-.0353	-.0220
ζ_2	0.0	-.1135	-.0358
φ (deg)	20.0	25.0	18.0
s (mm/s)	500.0	550.0	501.7
RMS		2641.0	2985.2
RESDIF			2648.3

parameter	2	3	4
p_1 (mm)	-499.1	-499.9	-499.9
p_2 (mm)	-202.5	-200.3	-200.0
p_3 (mm)	997.2	1000.0	1000.2
d_1 (mm)	993.2	999.6	999.9
d_2 (mm)	1218.8	1201.9	1200.2
d_3 (mm)	13.9	0.1	-1.0
ξ_1	.0031	-.0003	.0001
η_1	-1.000	-1.000	-1.000
ζ_1	-.0034	.0009	.0009
ξ_2	-1.000	-1.000	-1.000
η_2	-.0007	-.0001	-.0000
ζ_2	-.0098	-.0011	-0.000
φ (deg)	19.4	19.9	20.0
s (mm/s)	500.0	500.0	500.0
RMS	655.7	57.9	9.9
RESDIF	594.2	52.3	0.5

One additional iteration reduces the convergence margin RESDIF (i.e., the maximal weighted deviation between expected and obtained residuals) to $3.8 \cdot 10^{-5}$ without significantly changing the parameter estimates or the RMS. Thus in both examples the proposed method finds the correct solution within a few iterations with astounding accuracy, considering the noisy landmark positions used.

References:

[1] Gerald J. Bierman, *Factorization Methods for Discrete Sequential Estimation*, Academic Press, New York 1977

[2] Theodore D. Moyer, *Mathematical Formulation of the Double-Precision Orbit Determination Program (DPOPD)*, Technical Report 32-1527, Jet Propulsion Laboratory, Pasadena 1971

[3] Josef Stoer, *Einführung in die Numerische Mathematik I*, Springer-Verlag, Berlin – Heidelberg – New York ^{3}1979

Acknowledgment:

Financial support by Smiths Heimann GmbH Wiesbaden and the German Federal Ministry for Education and Research (under research grant no. 1763X07) is gratefully acknowledged.