ELSEVIER

Software alignment for tracking detectors

V. Blobel

Institut für Experimentalphysik, Universität Hamburg, Germany

Available online 14 June 2006

Abstract

Tracking detectors in high-energy physics experiments require an accurate determination of a large number of alignment parameters in order to allow a precise reconstruction of tracks and vertices. In addition to the initial optical survey and corrections for electronics and mechanical effects the use of tracks in a special software alignment is essential. A number of different methods is in use, ranging from simple residual-based procedures to complex fitting systems with many thousands of parameters. The methods are reviewed with respect to their mathematical basis and accuracy, and to aspects of the practical realization.

1. Introduction

Accurate alignment of tracking detectors is essential for important aspects of the physics analysis. The large and accurate vertex detectors of present and future experiments have a potential measurement precision of a few μm. The precision from mechanical mounting and e.g. LASER beam alignment is worse than the intrinsic resolution, and a high-precision alignment using tracks is required.

The general purpose of instrument calibration is explained in the statement: "Instrument calibration is intended to *eliminate or reduce bias* in an instrument's readings over a range for all continuous values. For this purpose, *reference standards* with known values for selected points covering the range of interest are measured with the instrument in question. Then a *functional relationship* is established between the values of the standards and the corresponding measurements" [1].

Software alignment/calibration of HEP track detectors is based, after the use of survey data and corrections for electronics and mechanical effects, mainly on track residual minimization. Real reference standards with known values do not exist and thus the alignment data may be incomplete, with several degrees of freedom undefined. Alignment/calibration requires to *understand* the detector (*functional relationship*) and to optimize thousands or tens of thousands of parameters. The goal is to reduce the χ^2 of the track fits, in order to improve track and vertex recognition, and to increase the precision of reconstructed tracks and vertices, eliminating or reducing bias in detector data.

2. Toy detector alignment

A popular alignment method in HEP is based on residual histograms. Tracks are fitted using a preliminary alignment. Histograms of hit residuals are generated and analyzed, and the offsets observed in the histograms are used to adjust the alignment. This method is tested below in a simple MC study.

A toy track detector is assumed with 10 drift chamber planes, 1 m high, 10 cm distance, no magnetic field, with a nominal accuracy of $\sigma \approx 200\,\mu m$ and an efficiency $\varepsilon = 90\%$ (plane 7 with reduced accuracy and efficiency). The chambers are displaced vertically by a certain shift of $\approx 0.1\,cm$. Ten thousand tracks with a total of 82 000 hits are generated and available for alignment.

The first alignment attempt is based on the distribution of hit residuals: a straight line is fitted to the track data,

E-mail addresses: volker.blobel@desy.de, blobel@mail.desy.de.
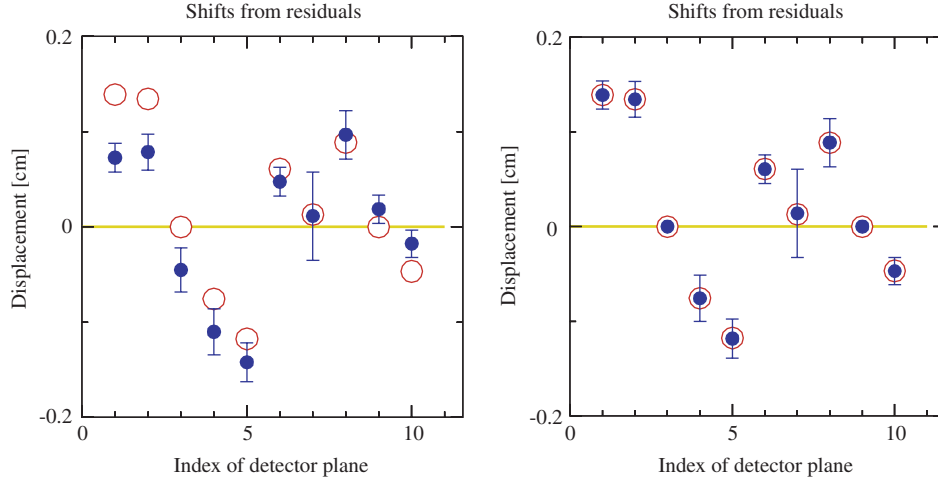*URL:* http://www.desy.de/~blobel.

Fig. 1. Alignment of the planes of a toy detector. The true shifts are given by the large circles. Results of the histogram alignment after 30 iterations are shown by points with error bars, on the left after the first attempt with 10 free parameters, and on the right with two parameters fixed at zero.

and the residuals (= measured vertical coordinate minus fitted coordinate) are histogrammed, separately for each plane. The mean value of the residuals is taken as correction to the vertical plane position, and the procedure is repeated iteratively. There are large changes in the first iteration, small changes in the second iteration, and almost no change afterwards. The result after 30 iterations is shown in Fig. 1 (left).

The reason for non-convergence to the true shifts is simple: two degrees of freedom are undefined, a simultaneous shift and an overall shearing of the planes. In a second attempt, the displacement of two planes (planes 3 and 9), assumed to be carefully aligned externally, is fixed at zero. After large changes in the first iteration there are smaller and smaller changes in the following iterations. The result after 30 iterations in Fig. 1 (right) shows that the method is converging, but rather slow, because the determination of displacements is based on biased fits.

Can the bias of the track-fit results due to the initially unaligned detector be avoided by a different method? A correct approach would be the simultaneous fit of the (global) alignment parameters $\boldsymbol{p}^{\text{global}}$ and all (local) track parameters $\boldsymbol{q}^{\text{local}}$, with the model

$$m_i \cong f(\boldsymbol{q}^{\text{local}}, \boldsymbol{p}^{\text{global}})_i = q_1^{\text{local}} + q_2^{\text{local}} \cdot s_i + \Delta p_j^{\text{global}}$$

for the measured value m_i, where $\Delta p_j^{\text{global}}$ is the shift for plane j. This is a linear least squares problem of $82\,000$ equations (measurements) and $20\,010$ parameters, which requires the solution of a linear equation with a $20\,010 \times 20\,010$ matrix. The special structure of the matrix, however, allows the reduction to an 8×8 matrix for the interesting plane-shift parameters, which are easily determined without iterations, if planes 3 and 9 are fixed at displacement $= 0$. This matrix reduction is explained in Section 4. The full covariance matrix is available after the fit and shows that the standard deviations of the shifts are around $3\,\mu\text{m}$. The method is easily extended to the determination of additional parameters e.g. the determination of corrections

$\Delta v_{\text{drift}}/v_{\text{drift}}$ of drift velocities for each plane, using the equations

$$\begin{aligned} m_i &\cong q_1^{\text{local}} + q_2^{\text{local}} \cdot s_i \\ &\quad + \Delta p_j^{\text{global}} + \ell_{\text{drift},i} \cdot \left(\frac{\Delta v_{\text{drift}}}{v_{\text{drift}}}\right)_j \end{aligned}$$

with a further reduction of the residuals. This improvement would be rather difficult to obtain with a pure residual-based method.

3. Special alignment methods

Histogramming: The basic idea of the histogram method, already mentioned in the previous section, is to extract parameter corrections from the peak (or mean or median) of residual histograms. The advantage of this method is that almost no extra code is necessary and the histograms can be generated from *n*-tuples of the residuals. However, the residuals are taken from biased fits and no precision alignment can be expected. In order to obtain convergence, many iterations would be necessary and therefore, the method is extremely slow. The method is limited to those parameters which are directly accessible from residuals histograms. Furthermore, it is not obvious how to fix undefined or badly defined degrees of freedom.

Residuals histograms can, however, be useful to detect large misalignments (see Fig. 2) of certain detectors, if also hits that are not assigned to the track are included in the histograms. The detection of very large local misalignments may be impossible otherwise.

Parametrization of residual dependence: An interesting method has been used for the internal alignment of the SLD vertex detector [2]. Starting from tracks reconstructed in the central drift chamber, different types of tracking constraints were classified in the three planes of the vertex detector. For each type of residuals, a functional form was derived and fitted to the measured residuals accumulated in
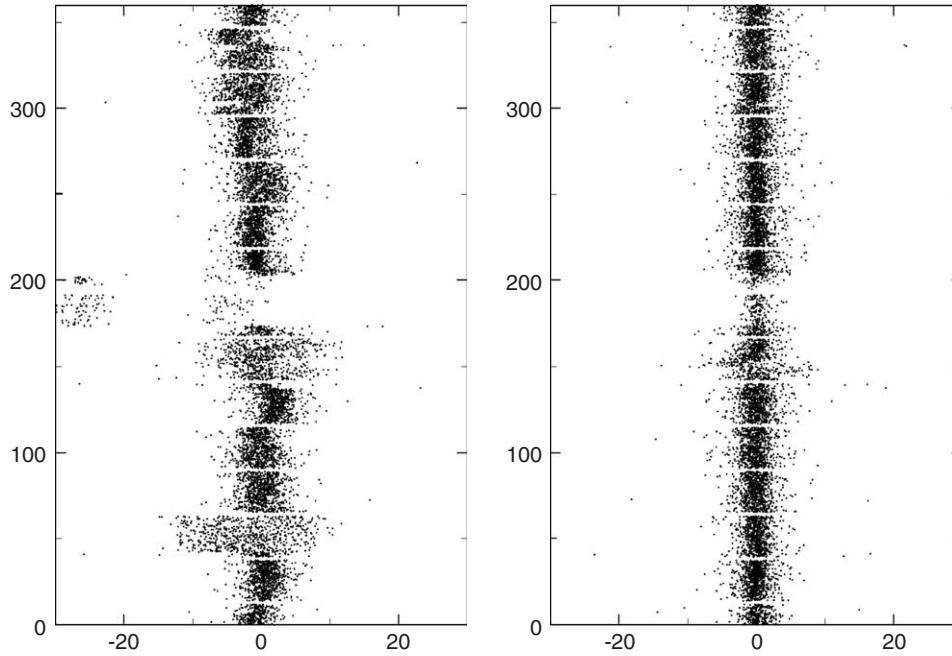
Fig. 2. Plot of residuals as a function of azimuthal angle before (left) and after (right) alignment.
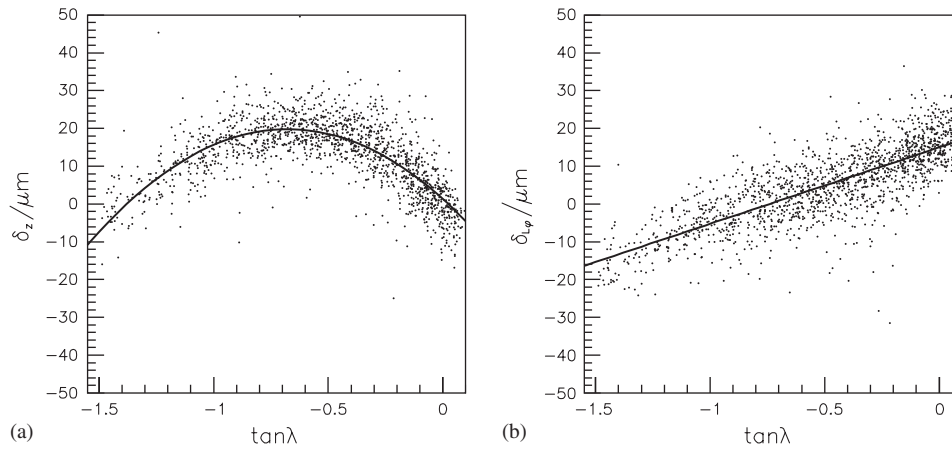
Fig. 3. Examples for residual fits in the SLD vertex detector alignment, here as a function of $\tan\lambda$ in a certain layer before the alignment [2].

n-tuples (Fig. 3). In total 2108 coefficients from 700 residual fits were determined. Taking into account in addition CCD shape corrections from the optical survey data of the CCD surfaces and taking into account the covariance matrices of the residual fits, 866 alignment corrections were determined from 5026 coefficients of the residual fits, using singular value decomposition (SVD) techniques for the least-squares fit minimization in a single step. With the aligned geometry a one-hit resolution of $4\,\mu m$ in both the rz and $r\phi$ planes was found, close to the true intrinsic CCD resolution, and the design performance was achieved. The post-alignment RMS of the residual distributions was found to be around a factor of four improved over the pre-alignment RMS values.

Alignment using a Kalman filter: A method for the estimation of alignment parameters during track reconstruction in parallel with the track parameter estimation using the Kalman filter has been studied [3]. After each track fit, the current alignment parameters are updated using an extension of the standard Kalman filter, which is a recursive least squares method.

Equations for updating the alignment parameters p and their covariance matrix V, as well as for the parameters of track k and their covariance matrix V_k are derived, which decouple into separate systems of equations. The correction to the alignment parameters is given by

$$\Delta p = V D^{\mathrm{T}} W (m - f).$$

The update requires the inversion of a matrix with the dimension of the measurement vector m as well as several matrix products:

$$W = (V_{\mathrm{meas}} + H V_k H^{\mathrm{T}} + D V D^{\mathrm{T}})^{-1}$$

where V_{meas} = covariance matrix of m, and D and H are the Jacobians of the track function f w.r.t. the alignment and track parameters.

The resulting alignment, considered to be a sort of *local* alignment, gives the positions and orientations of a set of detector elements with respect to a fixed set of reference detectors. Convergence has been studied in a MC simulation with six detectors to be aligned. Occasional convergence to local minima cannot be excluded; this problem is solved by the introduction of annealing, gradually stepping up the weights of the observations in the course of the estimation process.

4. Matrix methods for alignment

There are large correlations between the alignment parameters p, which specify the position and orientation of detector elements. The standard method for the determination of a large number of *n correlated* parameters is the solution of a system of linear equations, derived from a least squares sum of residuals, which has to be minimized. Corrections Δp are determined from the system of normal equations $C\Delta p = b$ with a symmetric $n \times n$ matrix C and a n-vector b. The problem is characterized not only by the large value of n, but also by the large amount of data from many events that have to be employed in the procedure in order to get accurate estimates for Δp.

4.1. Alignment parameters of a planar sensor

刚体有6个自由度

Six alignment parameters are required for a complete alignment of a planar sensor like a silicon pixel or strip detector. Coordinates and transformations are as follows [4]. Local (sensor) coordinates $q = (u, v, w)$, defined w.r.t. a sensor and used for track reconstruction, and the coordinates $r = (x, y, z)$ in the global detector system are related by the linear transformation

$$q = R(r - r_0)$$

with a nominal rotation matrix R, and the nominal position vector r_0. The origin of the (u, v, w) is at the center of the sensor; the u-axis along the precise coordinate and the v-axis along the coarse coordinate are in the sensor plane.

After alignment the transformation becomes

$$q_{aligned} = (R_\gamma R_\beta R_\alpha)R(r - r_0) - \Delta q$$

with the correction vector $\Delta q = (\Delta u, \Delta v, \Delta w)$; R_α, R_β and R_γ are rotation matrices, defined by (small) angles of rotation $\Delta\alpha$, $\Delta\beta$ and $\Delta\gamma$ around the u-axis, the (new) v-axis and the (new) w-axis. In the small-angle approximation the correction matrix for rotation becomes

$$R_\gamma R_\beta R_\alpha = \begin{pmatrix} 1 & \Delta\gamma & -\Delta\beta \\ -\Delta\gamma & 1 & \Delta\alpha \\ \Delta\beta & -\Delta\alpha & 1 \end{pmatrix}.$$

Usually not all parameters $(\Delta u, \Delta v, \Delta w)$ and $(\Delta\alpha, \Delta\beta, \Delta\gamma)$ are well-defined and it may be better to use a sub-set or selected linear combinations of the six parameters.

4.2. Global degrees of freedom

It is a trivial fact, that e.g. a global translation of the whole detector has no influence on the χ^2 of the track. Therefore, track residual minimization as the basic principle in detector alignment is not sufficient to fix all global degrees of freedom. Undefined or weakly defined global degrees of freedom may introduce certain distortions, which do not affect track-fit χ^2-values, but result in a bias of fitted track parameters. This has to be avoided.

A general linear transformation with a translation vector and a 3×3 matrix R

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} d_x \\ d_y \\ d_z \end{pmatrix} + R \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

depends on $3 + 9$ parameters; the matrix with 9 parameters can be decomposed into

- three rescaling factors f_x, f_y, f_z, of coordinate axes;
- three rotations $\mathscr{D}_x, \mathscr{D}_y, \mathscr{D}_z$ and
- three shearings $\mathscr{T}_{xz}, \mathscr{T}_{yz}, \mathscr{T}_{xy}$.

At least some of these linear-transformation parameters have to be fixed in an alignment procedure. In addition, there are weakly defined nonlinear transformations or deformations. Examples are: so-called *clocking*, i.e. a radius-dependent ϕ-shift, radial distortions, telescope effect by a radius-dependent z-shift, and sagitta effects [5]. Some of the effects can be reduced by the simultaneous use of different data sets in the alignment. The clocking effect will be reduced by using tracks with a vertex constraint. Mass-constraints in two-particle decays reduce sagitta effects. The telescope and sagitta effects are reduced or eliminated by the use of cosmics with magnetic field (large distance to interaction point) and without magnetic field (zero curvature). Cosmics and beam halo muons improve the alignment by introducing additional correlations between the alignment parameters. High-momentum tracks are preferred because of the larger predictive power compared to low-momentum tracks with large multiple scattering.

Alignment by tracks should be supplemented by the use of external information e.g. from survey and laser alignment. Fixing certain reference planes will improve the stability of the alignment fit. Undefined degrees of freedom can be avoided by adding equality constraint equations, e.g. *zero global displacement* in the x direction

$$d_x = \sum_i \Delta x_i = 0$$

or *zero rotation* of the whole detector. The standard technique is to introduce a Lagrange multiplier λ after linearization of the constraint $(g(p) + g^T \cdot \Delta p = 0)$ with vector $g = \partial g(p)/\partial p$, extending the matrix equation to

$$\left(\begin{array}{c|c} C^{\text{global}} & g \\ \hline g^T & 0 \end{array} \right) \left(\begin{array}{c} \Delta p^{\text{global}} \\ \hline \lambda \end{array} \right) = \left(\begin{array}{c} b^{\text{global}} \\ \hline -g(p) \end{array} \right).$$

Another method is to modify the solution of the matrix equation: singular value decomposition or diagonalization methods allow to recognize weakly defined linear combinations and to remove their effect (see Section 5).

4.3. Global minimization

Global alignment is based on the residuals obtained from the measurement of a large number of tracks. The alignment parameters p are global parameters. Parameters q_k for track k are local parameters. In the linear approximation the measurement equation for measurement m_i of the track k is written as

$$m_i \cong f(q^{\text{local}}, p^{\text{global}})_i + (\delta_i^{\text{local}})^T \Delta q_k^{\text{local}} + (d_i^{\text{global}})^T \Delta p^{\text{global}} \tag{1}$$

where $f(q^{\text{local}}, p^{\text{global}})_i$ is the track model prediction for m_i and d_i^{global} and δ_i^{local} are the derivative vectors w.r.t. the global and the local parameters. For a given alignment, a single track fit for track k is performed by χ^2-minimization of the weighted sum of residuals $r_i = m_i - f(q^{\text{local}}, p_k^{\text{global}})_i$ using the current alignment parameter values. The weight w_i is the inverse variance of the measurement m_i.

Single track fit: The normal equations of the least squares solution for the track parameters are

$$(\Gamma_k) \times (\Delta q_k) = (\beta_k) \tag{2}$$

with the symmetric matrix Γ_k and the vector β_k, given by the sums over all measurements of the track k:

$$\Gamma_k = \sum_i w_i \delta_i^{\text{local}} (\delta_i^{\text{local}})^T$$

$$\beta_k = \sum_i w_i r_i \delta_i^{\text{local}}.$$

Track parameter corrections are obtained iteratively by $\Delta q_k = \Gamma_k^{-1} \beta_k$ until the vector β_k becomes negligible. The covariance matrix of the track parameters after convergence is $V_k = \Gamma_k^{-1}$.

Simultaneous alignment and track fit: The so-called *global* χ^2-function is the least squares sum over a large number of tracks from different data sets

$$\chi^2 = \sum_{\text{data sets}} \left(\sum_{\text{events}} \left(\sum_{\text{tracks}} \left(\sum_{\text{hits}} w_i r_i^2 \right) \right) \right) \tag{3}$$

which is minimized either with respect to the alignment corrections Δp only (solution I), or with respect to the corrections Δp and *all* track parameters q_j (solution II).

The normal equations for the χ^2-function (3) of the alignment parameters and *all* track parameters are given by

$$\left(\begin{array}{c|ccc} C & \ldots & G_k & \ldots \\ \hline \ldots & \ldots & \ldots & \ldots \\ G_k^T & 0 & \Gamma_k & 0 \\ \ldots & \ldots & \ldots & \ldots \end{array} \right) \times \left(\begin{array}{c} \Delta p \\ \ldots \\ \Delta q_k \\ \ldots \end{array} \right) = \left(\begin{array}{c} b \\ \ldots \\ \beta_k \\ \ldots \end{array} \right) \tag{4}$$

where k is the track index. Submatrix C is a symmetric $n \times n$ matrix (n = number of alignment parameters) and there are $K \cdot m$ additional rows and columns, if K tracks each with m parameters contribute. The complete matrix of Eq. (4) is huge, and cannot be stored in memory.

In order to build-up the matrix of Eq. (4) single tracks are fitted in a loop over all events from all data sets. The $n \times n$ matrix C and the n-vector b of the (global) alignment parameters are formed, track by track, by a sum with contributions from all measurements; the update formulae are, for each track,

$$C := C + \sum_i w_i d_i^{\text{global}} (d_i^{\text{global}})^T \tag{5}$$

$$b := b + \sum_i w_i r_i d_i^{\text{global}}. \tag{6}$$

In addition, there is for each track k, a rectangular $n \times m$ matrix G_k, which correlates the parameters of track k with the alignment parameters. It is determined by a sum over all measurements of track k:

$$G_k = \sum_i w_i d_i^{\text{global}} (\delta_i^{\text{local}})^T.$$

The matrix equation (4) cannot be solved directly because of its size. Two methods for the determination of the corrections Δp of the alignment parameters are in use, which are described in the following.

Solution I: One possibility is to solve the huge system of linear equations (4) iteratively. First, all tracks are fitted using the current values of alignment parameters to form the sums for C and b and, ignoring the matrices G_k, the linear system

$$(C) \times (\Delta p) = (b) \tag{7}$$

is solved for alignment corrections Δp. This method only takes into account the direct correlation between the different alignment parameters, whereas the complete correlations between the different alignment parameters, mediated by the tracks, are neglected. Therefore, the single-track fits and the solution of Eq. (7) have to be repeated many times. Experience has shown that the iterative solution is converging, but many, perhaps several hundred, iterations are necessary to reach the final solution.

Applications: The method has been used successfully in several experiments, e.g. Ref. [6]. In the alignment of the upgraded vertex detector of ALEPH at LEP2 [7] a *global* χ^2

involving all $864 = 144 \times 6$ degrees of freedom is built, using single tracks and vertex constraints. Selected information is used from the outer tracking. Precise faces measurements are used to reduce the degrees of freedom, while allowing for parametrized distortions. The method is shown to provide accurate results, even with a limited number of events; 16 000 hadronic Z events are used.

A sensor alignment with repeated track fitting and residual optimization has been developed for the CMS experiment by Karimäki et al. [4]. As described in Section 4.1, six parameters are required for one planar sensor. Within solution I, there are no correlations between different sensors, and therefore the solution involves matrices whose dimension is at most 6×6. The method was applied in a precision survey in a test beam setup with several silicon strip detectors. The performance has also been tested in a simulated two-layer pixel detector, with 144 sensors in layer 1 and 240 sensors in layer 2, each with six parameters. Only a few iterations were necessary in the case where all sensors of the second layer were fixed. In the case of only a single fixed sensor and all remaining 383 sensors misaligned, the number of iterations needed to reach a reasonable precision for the 2298 fitted parameters varied between 20 and 100.

Solution II: A non-iterative determination of the alignment parameters is possible due to the special structure of the huge matrix in Eq. (4). As far as the alignment is concerned, Eq. (4) can be reduced to an equation for the alignment parameters Δp only, with a modified matrix C' and vector b'. This method has been developed for the MILLEPEDE package [8] for alignment.[1] As before, a loop over all tracks is performed with single track fits, and the contribution from each track is added:

$$C' := C' + \sum_i w_i d_i^{\text{global}} (d_i^{\text{global}})^{\text{T}} - G_k V_k G_k^T \qquad (8)$$

$$b' := b' + \sum_i w_i r_i d_i^{\text{global}} - G_k V_k \beta_k \qquad (9)$$

(usually the corrections to the vector b are zero, because the vectors β_k for tracks are zero). No further information from single tracks has to be stored. The matrix update includes the information from the parameters of track k. In particular, the complete correlations between the different alignment parameters are included. Finally, after accumulation of all information from all tracks the matrix equation

$$(C') \times (\Delta p) = (b') \qquad (10)$$

has to be solved. Matrices C' and vectors b' from several data sets can be simply added to get a combined result. The method allows to introduce equality constraints (see Section 4.2). The algorithm to solve large least squares problems in a single step (no iterations) and taking into account all correlations is general and can be applied to

other problems with a large number of (global) parameters and a huge number of measurements with local parameters. In the alignment, a few iterations may be necessary to check the code, because of linear approximations made in the track model, or because of the limited accuracy of the solution of the large system of linear equations. Another reason for iterations may be the removal of outliers, applying certain cuts which are tightened from iteration to iteration.

Applications: The MILLEPEDE package is used in several experiments (H1(1997) [9], CDF(2001) [10], HERA-B) or is under test for others (CMS, LHCb, PHENIX, ZEUS). Starting from a different ansatz, essentially the same method to reduce the matrix size was derived in the ATLAS alignment group (convenor A. Hicheur); the conference report [5] describes tests in simulated experiments and contains several formulae for vertex and mass constraints etc.

A common alignment and detailed calibration was performed in the H1 experiment using the MILLEPEDE package for the silicon detector (two planes) and the drift chambers (56 planes), with a total of 1400 parameters using 50 000 tracks [9]. For the two drift chambers, 14 global parameters representing an overall shift or tilt were introduced. Local variations of the drift velocity v_{drift} for cell halfs and layer halfs are observed and described by $180 + 112$ corrections, which change with the HV configuration. For each group of eight wires corrections to T_0 are introduced (330 corrections). The result is shown in Fig. 4, where the reduction of the mean track residuals from ≈ 180 to $\approx 125 \, \mu\text{m}$ is visible.

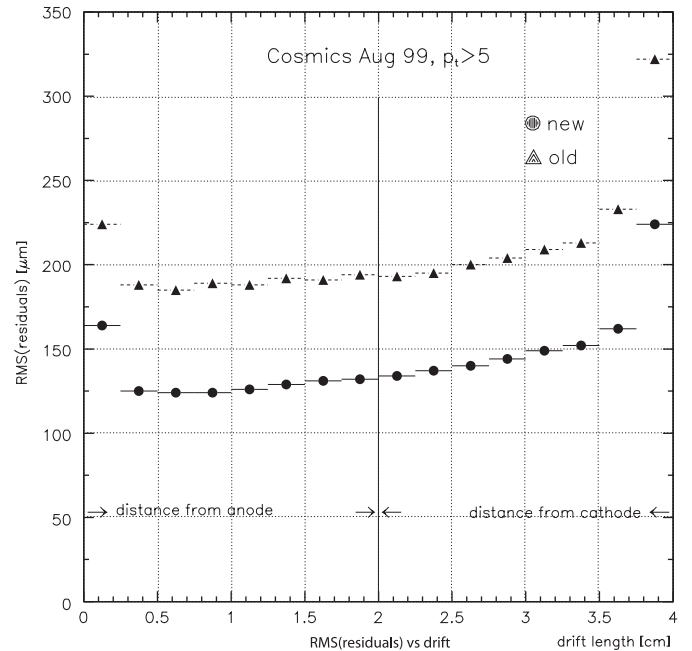
Fig. 4. Residuals as a function of the drift length in the H1 drift chamber. RMS values are reduced from 180 to 125 μm in the alignment with MILLEPEDE.

---

[1]There are indications that this matrix-size reduction in large least squares problems has been used in surveying already in the 19th century.

Because of CPU time and memory constraints the present MILLEPEDE package is limited to alignment problems with a number n of parameters up to five or ten thousand. A new version MILLEPEDE II is being developed applicable for n much larger than 10 000. In MILLEPEDE II, the two tasks of the program are split. Accumulation of data is done by a small subprogram (MILLE) inside the user program and the solution is determined in a stand-alone program (PEDE). An efficient sparse-matrix storage scheme is dynamically defined and different solution algorithms (see Section 5) are available. The solution is easily tested under different conditions using the accumulated data. Tests on a standard PC with 25 000 parameters from one million tracks took less than one hour.

5. Numerical linear algebra

In matrix methods the corrections Δp for alignment parameters are determined by the solution of matrix equations (7) and (10) with a large $n \times n$ matrix C. Double precision storage for matrix and vector is important already in the accumulation phase. It is not a priori clear whether a solution with acceptable accuracy can be obtained. The accuracy depends on the *algorithm* and on the *data*:

1. Algorithm: With a stable algorithm, the computed solution is the exact solution of a nearby problem. Gaussian elimination with restricted scan on the diagonal for the next pivot element is considered to be a stable algorithm for positive definite matrices.
2. Data: The system is called ill-conditioned, if small changes in the data can cause large changes in the solution. This behavior is caused by variables which are undefined or poorly defined or strongly correlated; all components of the solution are affected in this case. Ill-conditioning can be detected by small eigenvalues of the matrix and by global correlation coefficients close to 1. Re-definition of the alignment parameters may improve the condition of the system.

Matrix and vector should be appropriately scaled with consistent units in data and variables, in order to reach similar precision for all elements. A large fraction of the memory will often be used for the matrix; special storage techniques for symmetric matrices, band matrices and sparse matrices in general reduce the required space. Most algorithms can work *in-space*, i.e. no extra space is required for the inverted or decomposed matrix. Four solution methods are listed below.

Solution with matrix inversion: The standard method is the Gauss algorithm with pivot selection of the diagonal, with a computing time $\propto n^3$. Problems with $n =$ several thousand are solved on a standard PC within a time of the order of one hour. In practice, often at least a few parameters are badly defined (dead or inefficient channels)

and standard matrix routines will fail. A simple method to avoid such problems is to stop the inversion if no acceptable pivot is found, i.e. the largest possible submatrix is inverted with accurate results for the related parameters; the corrections for the remaining parameters are set to zero (subroutine SPINV in Millepede [8]).

An advantage of the method is that all variances and covariances are available with the inverse matrix, which is the covariance matrix for parameters: $V = C^{-1}$. The *global correlation coefficient* ρ_j

$$\rho_j = \sqrt{1 - \frac{1}{(V)_{jj} \cdot (C)_{jj}}}$$

can be calculated and gives a measure of the total amount of correlation between the jth parameter and *all* other variables. It is the largest correlation between the jth parameter and every possible linear combination of all the other variables and has a range from 0 to 1. Values of the global correlation coefficient close to 1 mean a large correlation and may indicate that too many partially redundant parameters were introduced. The accuracy of matrix inversion is reduced in this case.

Singular value decomposition and Diagonalization: These algorithms allow to recognize singularity or near-singularity of the matrix by the determination of singular values or eigenvalues, and this allows to ignore the corresponding linear combinations of parameters.

Diagonalization is the decomposition $C = UDU^{\mathrm{T}}$ with D diagonal (diagonal elements are the eigenvectors λ_j), and matrix U square and orthogonal with $UU^{\mathrm{T}} = U^{\mathrm{T}}U = 1$. The inverse is $C^{-1} = UD^{-1}U^{\mathrm{T}}$. The decomposition algorithms are iterative, with a computing time ≈ 10 times larger than for inversion, and additional space is needed for the $n \times n$ matrix U. The solution of $C\Delta p = b$ can be written in the form

$$\Delta p = U\left[\operatorname{diag}\left(\frac{1}{\lambda_i}\right)\right](U^{\mathrm{T}}b).$$

Insignificant linear combinations, which could produce distortions of the alignment, indicated by small eigenvalues, are suppressed by setting $1/\lambda_i = 0$ in the above formula for eigenvalues $\lambda_i = 0$ or small. This method is tested in the paper [5]; several tiny values appear in the spectrum of the eigenvalues, and those singular modes are suppressed to avoid potential distortions of the detector.

Another solution is to define a vector q by

$$q = \left[\operatorname{diag}\left(\frac{1}{\sqrt{\lambda_i}}\right)\right](U^{\mathrm{T}}b)$$

and to compute the solution Δp by

$$\Delta p = U\left[\operatorname{diag}\left(\frac{1}{\sqrt{\lambda_i}}\right)\right]q.$$

By construction, the vector q has a covariance matrix $V[q] = 1$, and this allows to recognize and suppress

insignificant contributions, recognized by a small value of $|q_i| \lesssim 1$.

Generalized minimal residual method (*GMRES*): The fraction of non-zero off-diagonal elements in the large matrix of a typical alignment problem is often rather small, of the order of a few percent. The approximate solution of a very large system of linear equations with a sparse matrix can be obtained by a certain algorithm for a quadratic minimization problem, in analogy to the method of conjugate gradients. One example is MINRES [11], designed to solve

$$\boldsymbol{C}\Delta\boldsymbol{p} = \boldsymbol{b} \quad \text{or} \quad \min \|\boldsymbol{C}\Delta\boldsymbol{p} - \boldsymbol{b}\|_2$$

where $\boldsymbol{C}$ is a symmetric matrix of logical size $n \times n$, which may be indefinite and/or singular, very large and sparse. The matrix is accessed *only* by means of a subroutine call which must return the product $\boldsymbol{y} = \boldsymbol{C}\boldsymbol{x}$ for any given vector $\boldsymbol{x}$. This iterative solution is faster by several orders of magnitudes compared to matrix inversion.

An example of a compact storage, optimized for the above product, is the row-index sparse storage [12] with two arrays of $(n + qn(n - 1)/2 + 2)$ words ($q =$ fraction of non-zero off-diagonal elements), one for real numbers and one for integers. 512 Mbytes of memory are sufficient for a matrix with $n = 100\,000$ for a value of q close to 1%.

Cholesky decomposition: The Cholesky decomposition $\boldsymbol{C} = \boldsymbol{L}\boldsymbol{D}\boldsymbol{L}^{\mathrm{T}}$ of the symmetric matrix $\boldsymbol{C}$ is numerically extremely stable, and can be made *in-space*; matrix $\boldsymbol{L}$ is a left unit triangular matrix (diagonal elements $= 1$) and $\boldsymbol{D}$ is a diagonal matrix. The solution of $\boldsymbol{C}\Delta\boldsymbol{p} = \boldsymbol{L}(\boldsymbol{D}\boldsymbol{L}^{\mathrm{T}}\Delta\boldsymbol{p}) = \boldsymbol{b}$ is obtained by forward and backward substitution. With a clever ordering of parameters the matrix $\boldsymbol{C}$ of alignment problems can be approximated by a *band* matrix. An important property of the Cholesky decomposition is the fact that for band matrices with band-width m the band structure is kept in this decomposition and the computing time is only $\propto m^2 \times n$. The subset of elements of the inverse matrix corresponding to the band of the original matrix $\boldsymbol{C}$ can be calculated quickly. Fast methods exist also for variable-bandwidth matrices (*sky-line* matrix), and for bordered band matrices (*arrow* matrix), where the border (additional full rows and columns) can for example be due to Lagrange multiplier constraints.

Several matrix algebra libraries exist, e.g. GNU Scientific Library (GSL), Numerical Algorithms Group (NAG), BLAS (Basic Linear Algebra Subprograms), and LA-PACK (Linear Algebra PACKage).

6. Alignment strategies and summary

Alignment problems with $n =$ several thousand parameters have been successfully solved, either using rather specialized methods, by algorithms requiring a large number of iterations or by global χ^2-minimization in a single step. The experience has shown that the integration of alignment into the reconstruction code and the use of fast alignment algorithms is of advantage; it allows a routine check of the time stability of alignment. The simultaneous use of several or all available types of events, physics *and* background events, and of single tracks and tracks with vertex and invariant-mass constraints can reduce or avoid potential distortions.

The next generation of experiments, e.g. the ATLAS and the CMS experiment at the LHC, have a huge number of independent sensors with an excellent spatial resolution from about $10\,\mu\mathrm{m}$ to about $50\,\mu\mathrm{m}$. An alignment precision for all sensors, with up to $n = 10^5$ parameters (CMS experiment at LHC), below the intrinsic resolution is required to get the necessary measurement accuracy for the physics program at the LHC.

The best and most suitable strategy for this large number of parameters is unknown at present. Both experiments have formed alignment working groups, where the impact of mis-alignment is studied [13] and different methods of alignment are developed in parallel and compared. The algorithms under study are rather similar in the two experiments. Both, simple straightforward and perhaps robust methods and advanced methods are studied in the ATLAS Alignment group (convenor A. Hicheur).

- Local χ^2 alignment: modules aligned on an individual basis with 6×6-matrices, iteratively using the alignment and refitting tracks.
- Global χ^2 alignment: simultaneous alignment and track fits with the full Pixel + SCT Barrel and Endcaps [5].
- Alignment with overlaps: relative module to module misalignment determined from overlap residuals.

as well as in the CMS Alignment group (coordinator O. Buchmüller)

- Sensor alignment by tracks: iterative procedure that considers individual measurement devices with 6×6-matrices (not taking into account correlations between measurement devices) [4,14],
- MILLEPEDE II: upgraded version of Millepede;
- Kalman filter alignment [3].

Some theoretical progress in alignment would be welcome to find a general strategy to suppress unwanted distortions, which can be used even for a very large value of n. However, there is confidence that good alignment strategies will be available at the time when they will be needed.

References

[1] National Institute of Standards and Technology, NIST/SEMATECH e-Handbook of Statistical Methods, 2005, ⟨www.itl.nist.gov/div898/handbook/⟩.
[2] D.J. Jackson, D. Su, F.J. Wickens, Nucl. Instr. and Meth. A 510 (2003) 233.

[3] R. Frühwirth, T. Todorov, M. Winkler, J. Phys. G: Nucl. Part. Phys. 29 (2003) 561.

[4] V. Karimäki, A. Heikkinen, T. Lampén, T. Lindén, Sensor alignment by tracks, CMS Conference Report CHEP03, La Jolla, CA, March 24–28, 2003.

[5] P. Brückman de Renstrom, S. Haywood, Least squares approach to the alignment of the generic tracking system, PHYSTAT2005, Oxford.

[6] A. Sopczak, Alignment of the D0 Vertex Detector, these proceedings.

[7] A. Bonissent, et al., Alignment of the upgraded VDET at LEP2, ALEPH 97-116, 1997.

[8] V. Blobel, Linear least squares fits with a large number of parameters, 2000, ⟨http://www.desy.de/~blobel⟩ including Fortran code.

[9] V. Blobel, C. Kleinwort, A new method for the high-precision alignment of track detectors, Proceedings PHYSTAT2002, Durham, arXiv-hep-ex/0208021.

[10] R. McNulty, et al., A procedure for the software alignment of the cdf silicon system, CDF/DOC/TRACKING/GROUP/5700, 2001.

[11] C.C. Paige, M.A. Saunders, SIAM J. Numer. Anal. 12 (4) (1975) 617.

[12] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipes—The Art of Scientific Computing, Cambridge University Press, Cambridge, 1999.

[13] N. De Filippis, Impact of CMS tracker misalignment on track and vertex reconstruction, these proceedings.

[14] T. Lampén, General alignment concept of CMS, these proceedings.