Correction Algorithms and High-Dimensional Characteristic Diagrams (B05)

Christian Naumann, Ilka Riedel, Ulrich Priber, Roland Herzog

Abstract The time-dependent consideration of the thermal situation in machine tools requires online-enabled methods for the correction of the tool position during the machining process. Higher-dimensional characteristic diagrams, which act as a lookup table for the prediction of the tool center point displacement from several temperature (or other) measurements, are well suited for this purpose. These characteristic diagrams are trained by finite element simulations. On the other hand, least-squares estimators together with reduced order models provide an alternative online-capable prediction technique for the tool displacement. In this context, optimal placement problems for the temperature sensors are described. In any case, the accuracy of simulations and estimation depends crucially on several parameters in the thermo-dynamic model. An approach that reveals the most relevant parameters using sensitivity analysis is also discussed.

1 Determination of Relevant Parameters Using Adjoint-Based Sensitivity Analysis

The accurate simulation of the thermo-elastic behaviour of machine tools requires good knowledge of model parameters and boundary conditions, see table 1. Some of these quantities, e.g., material parameters, are often well known, while others, such...
Table 1 Table of symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T$</td>
<td>temperature</td>
<td>K</td>
</tr>
<tr>
<td>$u$</td>
<td>displacement</td>
<td>m</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>stress</td>
<td>N/m$^2$</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>strain</td>
<td>1</td>
</tr>
<tr>
<td>$r$</td>
<td>thermal surface load</td>
<td>W/m$^2$</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>thermal conductivity</td>
<td>W/(Km)</td>
</tr>
<tr>
<td>$\rho$</td>
<td>density</td>
<td>kg/m$^3$</td>
</tr>
<tr>
<td>$c_p$</td>
<td>specific heat at constant pressure</td>
<td>J/(kgK)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>heat transfer coefficient</td>
<td>W/(Km$^2$)</td>
</tr>
<tr>
<td>$T_{\text{ref}}$</td>
<td>ambient temperature</td>
<td>$^\circ$C</td>
</tr>
<tr>
<td>$\nu$</td>
<td>Poisson’s ratio</td>
<td>1</td>
</tr>
<tr>
<td>$E$</td>
<td>Young’s modulus of elasticity</td>
<td>N/m$^2$</td>
</tr>
<tr>
<td>$\beta$</td>
<td>thermal volumetric expansion coefficient</td>
<td>1/K</td>
</tr>
</tbody>
</table>

as the heat transfer coefficient describing the thermal interaction between machine and environment, are more difficult to determine. Reliable values for the effective heat transfer coefficient are available in the literature only for simple geometric situations. For complex geometries, one has to resort either to computationally expensive CFD simulations of free convection/radiation problems, or model calibration using experimental data. Both approaches can be rather expensive.

It can be shown that the accurate determination of the effective heat transfer coefficient is not equally important in all parts of the machine structure. Knowledge about less important regions can help to significantly reduce experimental and computational efforts. The importance is based on how strongly inaccuracies in the parameter will cause inaccuracies in the value of the tool center point (TCP) displacement. An adjoint-based sensitivity analysis is proposed to identify regions of particular importance concerning the accurate value of model parameters. This technique can be used for any of the model parameters (or combinations thereof). It is elaborated below in the case of the heat transfer coefficient $\alpha$.

### 1.1 Background of Adjoint-Based Sensitivity Analysis

The thermo-elastic behaviour of a machine is modeled by a coupled system of partial differential equations. The first part of the system is the heat equation. Here the stationary case is considered,

$$\begin{align*}
- \text{div}(\lambda \nabla T) &= 0 \quad \text{in } \Omega, \quad \text{(1a)} \\
\lambda \frac{\partial T}{\partial n} + \alpha (T - T_{\text{ref}}) &= r \quad \text{on } \Gamma. \quad \text{(1b)}
\end{align*}$$
The domain $\Omega$ represents the geometry of the machine and $\Gamma$ is its boundary. The boundary conditions represent a simple model for the effective heat transfer which subsumes the transfer due to radiation and the free convection occurring at the lateral and top surfaces of the machine. The constant $T_{\text{ref}}$ denotes the ambient temperature.

The second part of the system is the elastic model, which consists of the balance of forces,

$$ -\text{div} \sigma = 0 \quad \text{in} \quad \Omega $$

involving the stress tensor $\sigma$. The relation between stresses and strains is given by Hooke’s law. An additive split of the linearized strain tensor $\varepsilon$ into its mechanically and thermally induced parts is employed, i.e.,

$$ \varepsilon = \varepsilon^{\text{el}} + \varepsilon^{\text{th}}, \quad \varepsilon^{\text{el}} = \mathcal{C}^{-1} \sigma := \frac{1 + \nu}{E} \sigma - \frac{\nu}{E} \text{trace}(\sigma) I, \quad \varepsilon^{\text{th}} = \beta (T - T_{\text{ref}}) I, $$

which yields an additive split of the stress tensor $\sigma$, i.e.,

$$ \sigma = \mathcal{C} \varepsilon - \beta (T - T_{\text{ref}}) \mathcal{C} I, \quad \mathcal{C} \tau = \frac{E}{1 + \nu} \tau + \frac{E \nu}{(1 + \nu)(1 - 2\nu)} \text{trace}(\tau) I. $$

The elastic model (2)–(3) is closed by the boundary conditions

$$ u = 0 \quad \text{on} \quad \Gamma_D, \quad (4a) $$

$$ \sigma \cdot n = 0 \quad \text{on} \quad \Gamma_N, \quad (4b) $$

which represent clamping on a part $\Gamma_D$ and absence of forces on the remainder of the surface $\Gamma$.

This coupled system (1)–(4) can be understood as an implicit equation

$$ e(x, \alpha) = 0 $$

with states $x = (T, u)$ consisting of the temperature and the displacement, and parameter $\alpha$.

In order to determine the sensitivity of the state w.r.t. the parameter, the calculation of the derivative of $x$ w.r.t. $\alpha$ is required. This is achieved by total differentiation of (5). The directional derivative (sensitivity) of $x \mapsto x(\alpha)$ at a nominal value $\alpha_0$ of the parameter and under variations in the direction $\delta \alpha$ is denoted by $x'(\alpha_0) \delta \alpha$. This quantity satisfies the following linear equation,

$$ e_x(x_0, \alpha_0) x'(\alpha_0) \delta \alpha + e_\alpha(x_0, \alpha_0) \delta \alpha = 0 $$

$$ \Rightarrow \quad x'(\alpha_0) \delta \alpha = -e_x(x_0, \alpha_0)^{-1} e_\alpha(x_0, \alpha_0) \delta \alpha. $$

Note that the sensitivity depends on the nominal value $(x_0, \alpha_0)$ for the states and the parameter due to the nonlinear coupling in the boundary condition (1b).
As mentioned above, one is not interested in the sensitivity of the entire state but only in the sensitivity of certain quantities such as the TCP displacement. Therefore, one additionally defines an output functional $f : X \to \mathbb{R}$, which maps the state onto the desired quantity. Then the directional sensitivity of the output w.r.t. changes in $\alpha$ follows from the chain rule,

$$\frac{d}{d\alpha} f(x(\alpha_0)) \delta \alpha = f'(x_0) x'(\alpha_0) \delta \alpha$$

$$= -f'(x_0) e_x(x_0, \alpha_0)^{-1} e_\alpha(x_0, \alpha_0) \delta \alpha \quad (6)$$

Evaluating (6) as is would require potentially repeatedly solving linearized thermo-mechanical models, as represented by the term $e_x(x_0, \alpha_0)^{-1}$ with multiple right hand sides (depending on the dimension of the parameter $\alpha$).

The main idea of the adjoint approach is to evaluate (6) in a different order by defining the so-called adjoint state $y$ through the adjoint equation,

$$e_x(x_0, \alpha_0)^\top y = -f'(x_0)^\top.\quad (7)$$

Then (6) can be rewritten as

$$\frac{d}{d\alpha} f(x(\alpha_0)) \delta \alpha = y^\top e_\alpha(x_0, \alpha_0) \delta \alpha. \quad (7)$$

It can be concluded that only one (adjoint) thermo-mechanical model has to be solved now. The approach is further illustrated by choosing $f(x) = u_z(TCP)$, i.e., the $z$-displacement of the TCP. Then the adjoint state $y = (S, v)$, consisting of adjoint temperature $S$ and adjoint displacement $v$, satisfies the following adjoint system:

$$-\text{div} (\lambda \nabla S) = \beta \frac{E}{1 - 2\nu} \text{trace}(\varepsilon(v)) \quad \text{in} \; \Omega,$$

$$-\text{div} \tau = -[0, 0, \delta_{TCP}]^\top \quad \text{in} \; \Omega$$

with the adjoint stress tensor $\tau = \mathcal{C} \varepsilon(v)$ and homogeneous boundary conditions

$$\lambda \frac{\partial S}{\partial n} + \alpha S = 0 \text{ on } \Gamma, \quad v = 0 \text{ on } \Gamma_D \quad \text{and} \quad \tau \cdot n = 0 \text{ on } \Gamma_N.$$

Having solved the adjoint system for $(S, v)$, the desired sensitivity can be efficiently calculated from (7), i.e.,

$$\frac{d}{d\alpha} f(x(\alpha_0)) \delta \alpha = \int_{\Gamma} \delta \alpha (T - T_{\text{ref}}) S \, ds \quad (8)$$

for arbitrary, in particular also local, perturbations $\delta \alpha$. 
1.2 Numerical Results

The method was tested for the Auerbach ACW630 machine column with two heat sources, one at the top of the motor flange and the other at the spindle nut. The picture shows the sensitivity factor \((T - T_{\text{ref}})S\) from (8) in [K mm/W]. In areas with high values the \(z\)-displacement at the TCP is very sensitive w.r.t. changes or inaccuracies in the effective heat transfer coefficient. Hence in these zones the heat transfer coefficient \(\alpha\) should be calculated (or determined experimentally) at high accuracy. On the other hand, in areas with low sensitivity values, inaccuracies in \(\alpha\) have only little influence on the \(z\)-displacement at the TCP.

2 Optimal Placement of Temperature Sensors for the Estimation of the TCP Displacement

The accuracy and robustness of an estimation of the TCP displacement from temperature measurements depends significantly on the locations where the thermal sensors are placed on the machine surface. This observation is true both for the estimation by means of characteristic diagrams, as in section 3, and for a direct estimation method, which is developed in the course of this section.

To be more precise, a simulation-based estimator is developed based on a reduced order model for the TCP displacement, which uses a number of temperature sensors spread across the machine surface. Later on, the sensor positions are optimized with the goal of increasing the robustness of the estimator, i.e., decreasing its sensitivity w.r.t. measurement errors. As in the previous section, numerical results are presented for the Auerbach ACW630 machine column.

2.1 TCP Displacement Estimation

With the aim of estimating the TCP displacement, the first goal is to reconstruct the entire temperature field from \(n_{\text{meas}}\) temperature measurements \(\hat{T}_j\), then calculate the displacement field and read out the desired quantity, i.e., the TCP displacement. The reconstruction of the temperature field is in general an ill-posed inverse problem. However, if the operating conditions, especially the heat inputs, of the machine are known, then a reduced order model can be set up a priori. Consequently, the
The temperature field can be represented as a linear combination of few basis vectors \( \varphi_i \in \mathbb{R}^{n_{\text{FE}}} \), also known as modes,

\[
T = T_{\text{ref}} + \sum_{i=1}^{n_{\text{POD}}} \zeta_i \varphi_i.
\]

These basis vectors, which can be interpreted as typical temperature distributions, are determined here by proper orthogonal decomposition (POD), see e.g. Kunisch and Volkwein (2001).

POD is a simulation-based model order reduction method. To apply it, one simulates the time-dependent heat equation,

\[
\rho c_p \dot{T} - \text{div}(\lambda T) = 0 \quad \text{in } \Omega,
\]

with boundary conditions (1b) and a possibly time-dependent heat source \( r(t) \), and saves snapshots \( \mathbb{R}^{n_{\text{FE}}} \ni y_j = T(t_j) - T_{\text{ref}} \) at a number of time points \( 0 = t_0 < t_1 < \ldots < t_{n_{\text{snap}}} \). The POD approach is to construct orthonormal basis vectors \( \varphi_i \) which approximate these snapshots as well as possible in the least-squares sense, i.e.

\[
\text{minimize } \sum_{j=1}^{n_{\text{snap}}} \left\| y_j - \sum_{i=1}^{n_{\text{POD}}} \langle y_j, \varphi_i \rangle_W \varphi_i \right\|_W^2, \quad \text{s.t. } \langle \varphi_i, \varphi_j \rangle_W = \delta_{ij}, \quad \varphi_1, \ldots, \varphi_{n_{\text{POD}}}. 
\]

Here \( W \) is a finite element mass matrix which represents a suitable inner product. The solution \( \varphi_1, \ldots, \varphi_{n_{\text{POD}}} \) of this minimization problem is given by the first \( n_{\text{POD}} \) eigenvectors of the generalized eigenvalue problem

\[
WYY^T \varphi_i = \lambda_i W \varphi_i,
\]

where \( Y = [y_1, \ldots, y_{n_{\text{snap}}} ] \) is the matrix of snapshots.

The reconstruction of the entire temperature field is now reduced to the determination of the few POD coefficients \( \zeta_1, \ldots, \zeta_{n_{\text{POD}}} \) such that the reconstructed temperature fits the measurements as well as possible in the least-squares sense,

\[
\text{minimize } \frac{1}{2} \sum_{j=3}^{n_{\text{meas}}} \left[ T_{\text{ref}} + \sum_{i=1}^{n_{\text{POD}}} \zeta_i \varphi_i(x_j) - \hat{T}_j(x_j) \right]^2, \quad \zeta \in \mathbb{R}^{n_{\text{POD}}}. 
\]

The solution of (11) is given by the normal equation

\[
J^\top J \zeta = J^\top (\hat{T} - T_{\text{ref}}),
\]

where \( \hat{T} - T_{\text{ref}} = (\hat{T}_1 - T_{\text{ref}}, \ldots, \hat{T}_{n_{\text{meas}}} - T_{\text{ref}}) \in \mathbb{R}^{n_{\text{meas}}} \) is the shifted vector of measurements and the Jacobian of the residual functions is given by
\[ J = J(x) = \begin{bmatrix}
\phi_1(x_1) & \cdots & \phi_{n_{POD}}(x_1) \\
\vdots & \ddots & \vdots \\
\phi_1(x_{n_{meas}}) & \cdots & \phi_{n_{POD}}(x_{n_{meas}})
\end{bmatrix} \in \mathbb{R}^{n_{meas} \times n_{POD}}. \]

Using (9), one can rewrite the reconstructed temperature as follows,
\[ T = T_{ref} + \Phi(J^\top J)^{-1}J^\top(\hat{T} - T_{ref}) \]
with \( \Phi = [\phi_1, \cdots, \phi_{n_{POD}}] \). The TCP displacement can then be calculated by first solving the elasticity equation (2)–(4), denoted by the operator \( S : T \mapsto u \),
\[ u = ST_{ref} + S\Phi(J^\top J)^{-1}J^\top(\hat{T} - T_{ref}) = 0 + S\Phi(J^\top J)^{-1}J^\top(\hat{T} - T_{ref}), \]
and then extracting the TCP components from the entire displacement field, i.e.,
\[ u(TCP) = B\Phi(J(x)^\top J(x))^{-1}J(x)^\top(\hat{T} - T_{ref}) := A(x) \]
with an observation matrix \( B \). Note that the matrix \( B\Phi \in \mathbb{R}^{3 \times n_{POD}} \) can be calculated offline by solving only \( n_{POD} \) elasticity equations and that during the operation of the estimator, no further solving of the heat equation nor the elasticity system is required. In other words, the reduced order POD model for the temperature naturally induces a reduced order model for the structural mechanics part as well.

The relation (12) can be understood as a linear estimator, which maps temperature measurements \( \hat{T} \) to the TCP displacement \( u(TCP) \). As the Jacobian \( J \) depends on the sensor positions \( x = [x_1, \cdots, x_{n_{meas}}] \), the reliability of the estimator depends on \( x \) as well.

### 2.2 Optimization of the Estimator’s Quality

It is well known that the reliability of a least-squares estimator, such as (12) can be judged by the eigenvalues of its covariance matrix. Assuming that all temperature sensors have the same standard deviation then the covariance matrix of the estimator is (up to a constant)
\[ C := \text{Cov}_{TCP}(x) = A(x)A(x)^\top = B\Phi(J(x)^\top J(x))^{-1}\Phi^\top S^\top B^\top. \]
Large eigenvalues in \( \text{Cov}_{TCP} \) point to a high sensitivity of the estimated TCP displacement w.r.t. perturbations in the temperature measurements, while small eigenvalues stand for a high robustness of the estimator against measurement errors.
This means that a good experimental setup, i.e., a good choice of sensor positions \( x \), is distinguished from a less suitable one by the eigenvalues of \( \text{Cov}_{TCP}(x) \). One possible criterion is the so-called E-criterion, which minimizes the largest eigenvalue of the covariance matrix. To this end, the optimal placement problem could be formulated as follows:

\[
\begin{align*}
\text{Minimize} & \quad \lambda_{\text{max}}(\text{Cov}_{TCP}(x)), \quad x = [x_1, \ldots, x_{n_{\text{meas}}}] \in \mathbb{R}^{3 \times n_{\text{meas}}} \\
\text{subject to} & \quad x_1, \ldots, x_{n_{\text{meas}}} \in \Gamma_{\text{feas}},
\end{align*}
\]

where \( \Gamma_{\text{feas}} \) are the parts of the machine surface where sensors can be placed.

The complexity of the optimization problem (14) is based on at least two distinct features. On the one hand, the eigenvalues of \( \text{Cov}_{TCP} \) are a continuous but non-differentiable function w.r.t. the location matrix \( x \), which would require specialized optimization solvers. On the other hand, it would be necessary to parametrize the constraint \( x_1, \ldots, x_{n_{\text{meas}}} \in \Gamma_{\text{feas}} \) with a complicated surface such as that of a machine column. In the sequel, the complexity of the problem will be reduced to make it tractable.

First, sensor locations are allowed only in a finite, possibly large, subset \( \Gamma_{\text{finite}} \) of \( \Gamma_{\text{feas}} \). An obvious choice for \( \Gamma_{\text{finite}} \) would be the nodes of the finite element mesh. The placement problem then becomes a high dimensional combinatorial optimization problem. For the machine column it is \( |\Gamma_{\text{finite}}| = 25,288 \) and there would be \( 319,728,828 \) possibilities to place only two sensors. To this end, the simultaneous placement of all sensors as in (14) is replaced by a sequential (greedy) placement of one sensor at a time, see algorithm 1. This will lead to sub-optimal solutions of (14), but render the overall problem tractable. Numerical evidence in Herzog and Riedel (2013) justifies this approach.

**Algorithm 1 (Sequential placement)**

1: Set \( i := 0 \)
2: repeat
3: \( i := i + 1 \)
4: Solve problem (15)
5: until \( i \geq n_{\text{POD}} \) and stopping criterion (17) is met

In each step of the sequential placement the position of one sensor is determined by solving the optimization problem (15), while other sensors at positions \( x_1, \ldots, x_{i-1} \) have already been placed and will be kept fixed.

\[
\begin{align*}
\text{Minimize} & \quad \lambda_{\text{max}}(\text{Cov}_{TCP}(x_i)) \\
\text{subject to} & \quad x_i \in \Gamma_{\text{feas}}.
\end{align*}
\]

Herein,

\[
\text{Cov}_{TCP}(x_i) = BS\Phi_i(J_i(x_i)^\top J_i(x_i))^{-1}\Phi_i^\top S^\top B^\top
\]
denotes the covariance matrix of the estimator in stage \(i\), which maps the first \(i\) temperature measurements to the TCP displacement vector. Herein, \(J_i\) and \(\Phi_i\) are restrictions of \(J\) and \(\Phi\), see Herzog and Riedel (2013) for details.

Besides making the placement tractable, the sequential placement approach allows the determination of the number of sensors \(n_{\text{meas}}\) on the fly rather than a priori. Algorithm 1 is stopped as soon as \(i \geq n_{\text{POD}}\) and

\[
\lambda_{\text{max}}(\text{COV}_{\text{TCP}}(x_i)) \leq \left( \frac{\epsilon_{\text{tol}}}{\sigma r(\alpha)} \right)^2
\]

with a user-defined tolerance \(\epsilon_{\text{tol}}\), the standard deviation \(\sigma\) of the sensors, a confidence level \(\alpha \in [0, 1]\) and the squared radius \(r(\alpha)^2 = F^{-1}_3(\alpha)\) the inverse cumulative distribution function of a \(\chi^2\) distribution with three degrees of freedom. The condition (17) guarantees that the largest half axis of the confidence ellipsoid describing the estimated tool center point displacement will be at most \(\epsilon_{\text{tol}}\).

2.3 Numerical Results

The sensor placement was tested again for the Auerbach ACW630 machine column. Two heat sources were assumed, one on top of the motor flange and the other at the spindle nut, see figure 1(a). The TCP displacement was approximated by the mean displacement of the mounting points shown in subfigure 1(b) of the sledge holding the spindle with the tool, shown in figure 1(b).

A POD simulation was carried out by heating each of the two sources until a stationary state was reached and cooling it down to the ambient temperature afterwards. It turned out that \(n_{\text{POD}} = 5\) basis vectors, figure 1(c) shows the first three modes, yield a pretty good reduced model for the temperature.

For the constants in the stopping criterion (17) of algorithm 1 a standard deviation of \(3\sigma = 0.1\) K was chosen, i.e., 99.73% of all temperature measurements lie within \(\pm 0.1\) K of their true value, a confidence level of \(\alpha = 0.95\) and a tolerance \(\epsilon_{\text{tol}} = 1 \cdot 10^{-6}\) m. Then one obtains that \(n_{\text{meas}} = 6\) sensors are enough to reach the desired accuracy of \(\epsilon_{\text{tol}}\). Figure 2 shows the square root of the maximal eigenvalue of the covariance matrix (16) during the first four placement problems, as well as the optimal sensor locations (yellow circles) at the points where the largest eigenvalue is minimal. As expected, all sensors lie near the two heat sources at the motor flange and the spindle nut.

Placing sensors in this way, one obtains a Jacobian \(J(x) \in \mathbb{R}^{n_{\text{meas}} \times n_{\text{POD}}} = \mathbb{R}^{6 \times 5}\), which leads (in a suboptimal sense due to the sequential placement) to a TCP displacement estimator (12) with minimal variance in the sense of (14).
3 Characteristic Diagrams

An alternative way of estimating the TCP displacement from temperature measurements according to (12) is through the use of characteristic diagrams. This can be conceived as a multi-dimensional lookup table which has been setup and trained in an offline phase. Typical input dimensions include temperature measurements at certain locations, but might include other information, such as the current geometrical configuration, such as the current spindle height, as well. In this respect, characteristic diagrams are more flexible than direct estimators such as (12). The output in the present context would consist of the three components of the TCP displacement.
Effectively, the use of a characteristic diagram corresponds to an interpolation and extrapolation of thermal states at selected locations (or selected grid points in a FE mesh) which participated in the training phase. Depending on the machine geometry and the FE discretization used, there will be many geometrical grid points with very little significance to the TCP displacement, either because their temperature is highly dependent on that of their neighbors, or because deformations in their vicinity have little effect on the TCP. Therefore, a few grid points are expected to be sufficient to estimate the TCP displacement with good accuracy.

Finding out which grid points matter most and how many are at least needed to achieve a given accuracy is subject of the sensitivity analysis and can be achieved with the methods described in section 2. For now, it is simply assumed that a small number \( n_g \) of such grid points suffices for the prediction. By determining the maximum and minimum temperature each grid point may reach during the machine tool’s operation and splitting the resulting intervals into discrete segments one gets \( n_g \) 1D grids, which are combined into one large \( n_g \)-dimensional grid of temperature variables. Each point in this grid represents one thermal state to which a TCP displacement vector is assigned.

\[
f : t = (T_1, T_2, \ldots, T_{n_g}) \in \mathbb{R}^{n_g} \rightarrow (x, y, z)_{TCP} \in \mathbb{R}^3
\]

Using data gathered in experiments or obtained through simulation, an attempt was made to fit an \( n_g \)-dimensional hypersurface into this \((n_g + 1)\)-dimensional space, see figure 3. As with any fitting, these samples need to cover as much of the grid as possible. Especially with measured data points one also needs some redundancy to reduce the effect of random errors.

A method called Smoothed Grid Regression (SGR) that was developed prior to this project, see Priber (2003) and Naumann (2012), uses least squares to calculate such hypersurfaces for a given temperature grid and a database of samples. It does so by weighting the penalties resulting from data errors, i.e., the distance of the samples from the hypersurface, against those resulting from a lack of smoothness, i.e., from large changes in the deformation gradient. This will lead to a large and
over-determined linear system such as

\[
\begin{pmatrix}
  z \\
  0
\end{pmatrix} = \begin{pmatrix}
  A \\
  S
\end{pmatrix} R,
\]

where each column of \( A \) and \( S \) corresponds to a grid vertex. Each row of matrix \( A \) describes the location of a data point w.r.t. its surrounding grid points while its value is the corresponding entry in \( z \). Each row of matrix \( S \) requires a grid point’s value to be colinear to its left and right neighboring values in every grid dimension.

While the SGR method requires very little knowledge of the underlying problem, its approximation quality heavily depends on the choice of input variables used for the grid. Another weakness of this method is the exponential growth of the resulting linear system with the dimension of the grid. Even though the assembled matrices have a very sparse structure, with only few elements in each row and column, problems with memory shortage arise quickly as the grid dimension increases. If direct solvers are used to solve these linear systems (in a least-squares sense) then the unavoidable fill-in will cause an overflow much sooner still. The same goes for the computation time, which grows along with the grid size. The main goal is therefore to make very large least-squares problems computable in order to boost the approximation quality of characteristic diagrams.

To solve these performance issues three elements are needed: a fast iterative solver, a way of solving the linear system without having to assemble the entire matrix and a way of reducing a large complete grid to a much smaller grid without much loss of accuracy. While there are a number of fast direct solvers, some of which make very good use of the SGR’s band matrix structure, they are bound to create a certain fill-in which grows rapidly for an increasing grid fineness. They also have trouble using the underlying grid and with assembling the matrix piece-by-piece during the solution process. For all these reasons iterative solvers become attractive alternatives. Since the efficiency of iterative solvers for symmetric problems (such as least-squares) largely depends on the condition number and distribution of eigenvalues, preconditioning becomes an important aspect. The SGR is a grid based method where the fineness of the grid can be arbitrary. Therefore, it is easy to establish a set of hierarchical grids that can be used by a multigrid solver or a multigrid preconditioner, e.g., in a conjugate gradient method.
Multigrid methods, see e.g. Trottenberg et al (2001), were originally developed to solve linear systems arising through the discretization of a partial differential equation (PDE), e.g., from a finite element approach. By expressing the regression problem as an energy minimization problem

\[
\min_u \frac{1}{2} \sum_{i=1}^{n_{\text{data}}} |z_i - u(x_i)|^2 + \frac{1}{2} \int_{\Omega} \lambda(x) \cdot |\nabla u(x)|^2 \, dx,
\]

one gets the following PDE in weak form:

\[
\forall \delta u : \sum_{i=1}^{n_{\text{data}}} (z_i - u(x_i)) \cdot \delta u(x_i) + \int_{\Omega} \lambda(x) \cdot \nabla u(x) \cdot \nabla \delta u(x) \, dx = 0.
\]

The second term in the energy functional corresponds to a smoothing term. The one shown above prefers constant values of the unknown function \(u\), whereas terms favoring constant gradients are also used frequently.

Experimentally it was shown that this SGR-based multigrid preconditioned CG solver achieves optimal complexity. Even the best iterative solver cannot beat \(O(n_p)\) iterations, where \(n_p\) is the total number of vertices in the temperature grid. It is thus essential to make the grid as small as possible for a given, fixed accuracy. Using an orthogonal grid is quite useful both in the multigrid and in the FEM but a simple grid in which every grid axis corresponds to one geometrical node leads to a hypercube of thermal states in which large parts can never be realized during the machine tool’s operation. This problem calls for a principal component analysis of the input variables (temperatures). By rotating the hypercube and shrinking it to form a kind of minimal orthogonal convex hull around the relevant thermal states, one accomplishes the task of concentrating most of the information in only a few wide dimensions, which makes it possible to cut off all the very narrow dimensions and to use a coarser discretization on the intermediate ones.

Another technique for reducing the number of grid points involves the use of adaptive FEM. Rather than discretizing the entire hypercube uniformly down to the finest level, one can also make the discretization finer in those regions with the greatest data error. In this particular case these are the states with large second derivatives of the TCP displacement \(u(x)\).

In the following test performed on experimentally measured data, the results of the described characteristic diagram approach are demonstrated. To do this a training data set and a test data set were assembled from two measurements on a machine tool. Figure 4 shows how well the characteristic diagram based on the training data predicts the displacements of the test data.

One can see that the predicted TCP displacement (thin, black) roughly follows the actual displacement (thick, gray). While the training data (not shown here) can be predicted with no more than \(\pm 10 \mu m\), the test data still produces a maximum error of \(\pm 20 \mu m\). This reduces the maximum error to one third, which is a good start. Parts of the predicted curve that strongly deviate from the measured curve suggest that the test data contradicts the training data. This most likely means that some data points
that lie in the same region of the grid strongly differ in a dimension that cannot be seen. One can therefore not fully distinguish all relevant thermal states of the machine tool from this thermal sensor setup, which shows the importance of the choice of input variables for the quality of any characteristic diagram. Similar tests have shown that a poor choice of input variables leads to particularly large errors when the temperature gradient is big. In these cases some machine tool parts have already heated up, expanded and thus caused a TCP displacement before any of the thermal sensors have picked it up. The opposite case has also occurred, where the thermal sensors have picked up heat before any of the relevant machine parts have started expanding.

Besides further advances in the efficient calculation of characteristic diagrams, another interesting research topic will be the improvement of predictions with a non-ideal choice of input variables. This will be particularly relevant in cases in which the sensor placement is restricted by safety or practical limitations.

4 Integration into the SFB/TR-96 and Outlook

To summarize, project B05 is one out of three projects dealing with real-time correction algorithms, where the time scale is determined by the thermal constants of the machine tool. All three approaches are based on a prediction of the thermally induced TCP displacement in order to allow for its correction through the machine’s NC controller.

The three approaches differ with respect to the input data that serve as the basis of the TCP displacement prediction. They also differ with respect to the way in which the thermo-mechanical functional chain is taken into account. Project B05 offers two algorithms to achieve the prediction of the TCP displacement. First, characteristic diagrams (Section 3) capture the thermo-mechanical behavior through a multidimensional lookup table which needs to be setup in an offline phase either by
measurements or simulated data. Second, a least-squares estimator (Section 2) has been developed that incorporates a reduced order model of the thermo-mechanical effects. In contrast to B06 and B07, both algorithms typically use temperature measurements, which are easy to gather, as their input, but are open to extension to other input data. Since the temperature data constitute part of the state of the machine (in contrast to, e.g., thermal loads), both algorithms in B05 avoid a drift-off of the predicted vs. the true thermal state due to model inaccuracies. A final distinguishing feature of the approach taken in project B05 is the optimal placement of temperature sensors according to their information content with regard to the TCP displacement. This helps to reduce the amount of input data and further adds to the robustness of the estimator.

It should be mentioned that all results so far were obtained for a static pose of the machine tool and it is ongoing work to incorporate changes in the pose. Moreover, we emphasize that all projects that employ simulation technology, among them A05, A06, A07, B05, B07, rely on the accuracy of model parameters for realistic results. Most notably, this concerns the heat transfer coefficient, which determines the rate of heat exchange with the machine’s environment. As shown in Section 1, however, the TCP displacement may depend strongly on the actual value(s) of this parameter in certain regions of the surface. Therefore, the real-time identification of the heat transfer coefficient will be a major goal for the future work in this project.

References