On conformal geospatial transformations with complex polynomials

Jüri Lippusa* and Tõnu Ojab

* Institute of Cybernetics, Tallinn University of Technology, Akadeemia tee 21, 12618 Tallinn, Estonia
b Institute of Ecology and Earth Sciences, University of Tartu, Vanemuise 46, 51014 Tartu, Estonia; tonu.oja@ut.ee

Received 28 March 2011, revised 8 November 2011, accepted 2 January 2012, available online 15 February 2012

Abstract. We consider conformal reprojections of geographic information system data with complex polynomials as compared to reprojections with commercial software packages. We show that in regular cases low-degree complex polynomials can be effectively used instead of sophisticated software packages. In less regular cases the ambiguities arising from other sources make the application of standard transforms problematic.

Key words: cartography, conformal mappings, complex polynomials.

1. INTRODUCTION

When talking about conformal mappings of a planar region onto another planar region a mathematician usually first thinks about complex analysis, the Riemann mapping theorem, and mapping with analytic functions. A geographic information system (GIS) specialist first thinks about the catalogue of conformal map projections, unprojection, geographic datum transformation, and reprojection with various software packages.

Another aspect that has to be considered is that conformal map projections often include quite complicated mathematics. The fact that the complexity is hidden inside software packages does not make it non-existent. The packages usually employ different approximation methods and depending on the quality of the approximation, the numerical results may vary.

The purpose of the present paper is to demonstrate that if our data in a limited area can be transformed with acceptable accuracy, using unprojection, geographic transformation, and reprojection, they can be transformed with comparable accuracy, using low-degree complex polynomials. This makes conflation of geographic data from different sources significantly faster and more transparent.

We consider some conformal map projections in the common form as presented by Snyder (1987) as our reference basis. These forms are common in many textbooks on geodesy and give a millimetre accuracy in the normal area of application. More precise formulas are available for many projections (e.g. Bugayevskiy and Snyder, 1995; Kuittinen et al., 2006, Appendix 1; Karney, 2011), but for our purpose millimetre accuracy of the formulas is sufficient.

Datums can be thought about in two ways: geometrical and statistical. In the geometrical approach datum means the location and orientation of the reference ellipsoid in space. This naturally leads to considering the 7-parameter (3-parameter, 9-parameter) transforms (Helmert, „Buša–Wolf, Molodensky, Molodensky–Badekas, etc.). In the statistical approach datum means a catalogue of coordinates of reference points. This in turn leads to various regression methods (multiple regression equations, triangulation, grid shift, etc.). A nice comparison of the two approaches can be found in Ordnance Survey U.K. (2008, Section 3.2).

A large number of papers have been devoted to datum transformations: Vaníček and Steeves (1996), Soler (1998), Iz and Chen (1999), to name just a
random few. A thorough systematic presentation of the geometrical approach can be found in Awange et al. (2010). An interesting paper by Chen and Hill (2005) discusses various quality issues of both approaches.

For reference we also consider three common commercial software packages. Let us denote them PackageA, PackageB, and PackageC.

Our test area is a rectangle covering the territory of Estonia between approximately 56.05°–61.13° N and 21.64°–28.58° E. This is the area for which we also have reference data from other sources available. We took a uniform grid with the step of 25 km in both directions, using the coordinate system of the Estonian National Grid L-EST97 (EPSG:3301). This gave us 187 test points. The actual calculations with the formulas from Snyder (1987) were performed with the computer package Mathematica®.

For numerical values of all parameters we used European Petroleum Survey Group (EPSG) Geodetic Parameter Registry Version: 7.6.3 (http://www.epsg-registry.org/). This registry is currently maintained by the Geomatics Committee of the International Association of Oil and Gas Producers and is the definitive data source for many commercial software packages and web mapping applications. The standard way to identify a coordinate system in web mapping applications is by its code in the EPSG registry.

Some abbreviations commonly used throughout the paper:

WGS 84 – World Geodetic System 1984. This is the reference system used by the Global Positioning System (GPS);
MSE – mean square error;
EST97 – a synonym for the Estonian National Grid L-EST97 that is used in the EPSG registry.

2. PRELIMINARIES

Let \( \mathbb{C} \) denote the set of all complex numbers \( z = x + iy \), where \( i \) is the imaginary unit. This set is also called the complex plane.

We present here some well-known results from complex analysis. We give the references following the book by Krantz (1999), but they can be found in many textbooks on complex analysis.

- Let \( D \) be a connected open subset of the complex plane \( \mathbb{C} \). Let \( f, g \) be holomorphic on \( D \). If \( S = \{ z \in D : f(z) = g(z) \} \) has an accumulation point in \( D \), then \( f \equiv g \) (Krantz, 1999, Corollary 3.2.3.1, p. 38). In particular this holds when \( S \) is an open subset of \( D \).
- The Riemann Mapping Theorem. If \( D \) is a non-empty simply connected open subset of the complex plane \( \mathbb{C} \) which is not all of \( \mathbb{C} \), then there exists a biholomorphic (bijective and holomorphic) mapping \( f \) from \( D \) onto the open unit disk \( U = \{ z \in \mathbb{C} : |z| < 1 \} \) (Krantz, 1999, Section 6.4.3, p. 87).
- If a non-empty simply connected open subset \( D \) of the complex plane \( \mathbb{C} \) is isomorphic to the unit disc \( U \), then the family of all conformal mappings of \( D \) onto \( U \) depends on three real-valued parameters. In particular, there exists a unique conformal mapping \( f \) of \( D \) onto \( U \) that is normalized by the conditions

\[
f(z_0) = 0, \quad \arg f'(z_0) = \theta,
\]

where \( z_0 \) is an arbitrary point in \( D \) and \( \theta \) is an arbitrary real number (Krantz, 1999, Section 6.2.1, p. 80).

The important conclusion from these results is that if a regular planar region (e.g. a rectangle) is conformally mapped onto another regular region, then this mapping is performed by a holomorphic function. This mapping is fully determined by its values on an arbitrarily small open subset of the region. A natural approximation of a holomorphic function is a complex polynomial.

3. DIFFERENT PROJECTIONS, IDENTICAL DATUMS

Let us consider conformal projections. The two projections that are of most interest in our test area are the Lambert conformal conic projection with two standard parallels, which is used in the coordinate system L-EST97 (EPSG:3301), and Transverse Mercator with central meridian 24°, which is used in the coordinate system TM-Baltic93 (EPSG:25884). Both coordinate systems have the same datum – European Terrestrial Reference System 1989 (ETRS89). These projections have no singular points in our area of interest, therefore in view of Section 2 there exists an analytic function that maps \( D \) onto \( D' \). We can approximate this function with a complex polynomial of appropriate degree. The degree depends on the required accuracy of the approximation.

The results are gathered in Table 1. In the column MSE the mean square differences and in the column Maximal the maximal differences and in the column MSE the mean square differences from the reference basis are presented. The reference basis, as mentioned above, is calculated, using the formulas from Snyder (1987). The distribution of residuals for the polynomials of degree 5, corresponding to the third row of the table, is shown in Fig. 1.
Table 1. Differences (m) for mappings from L-EST97 into TM-Baltic93

<table>
<thead>
<tr>
<th>Method</th>
<th>Maximal</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polynomials 3</td>
<td>0.6582</td>
<td>0.2095</td>
</tr>
<tr>
<td>Polynomials 4</td>
<td>0.0079</td>
<td>0.0022</td>
</tr>
<tr>
<td>Polynomials 5</td>
<td>0.0004</td>
<td>0.0001</td>
</tr>
<tr>
<td>PackageA</td>
<td>0.0086</td>
<td>0.0045</td>
</tr>
<tr>
<td>PackageB</td>
<td>0.1104</td>
<td>0.0706</td>
</tr>
<tr>
<td>PackageC</td>
<td>0.0007</td>
<td>0.0007</td>
</tr>
</tbody>
</table>

Fig. 1. Distribution of residuals, TM-Baltic93 approximated with polynomials of degree 5. The length of the longest arrow does not exceed 0.4 mm.

Another interesting case is transformation from L-EST97 into World Mercator WGS 84 (EPSG:3395). Our interest is caused by two aspects:
1. Up to a constant factor the coordinates are the same as conformal latitude and longitude (see König and Weise, 1951, Chapter 2, Section 2).
2. This projection is used by Google Maps and Google Earth (see, e.g., Google, 2011). The corresponding results are presented in Table 2 and Fig. 2.

The degree of the polynomials that are needed in this case is slightly higher. This is caused by the fact that in the first case the projections were specially chosen so that our area of interest would lie in the area of small distortions.

Analogous results can be obtained for all conformal projections where there are no singular points in the area of interest. When the area of interest is in the region of small distortions, complex polynomials of relatively low degree are sufficient to produce good approximations. When the area of interest is in the area of large distortions, we have to raise the degree of the polynomial.

Table 2. Differences (m) for mappings from L-EST97 into World Mercator WGS 84

<table>
<thead>
<tr>
<th>Method</th>
<th>Max</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polynomials 4</td>
<td>0.8668</td>
<td>0.2491</td>
</tr>
<tr>
<td>Polynomials 5</td>
<td>0.0367</td>
<td>0.0100</td>
</tr>
<tr>
<td>Polynomials 6</td>
<td>0.0016</td>
<td>0.0004</td>
</tr>
<tr>
<td>PackageA</td>
<td>0.0029</td>
<td>0.0028</td>
</tr>
<tr>
<td>PackageB</td>
<td>0.2138</td>
<td>0.1363</td>
</tr>
<tr>
<td>PackageC</td>
<td>0.0002</td>
<td>0.0002</td>
</tr>
</tbody>
</table>

Fig. 2. Distribution of residuals, World Mercator approximated with polynomials of degree 5. The length of the longest arrow does not exceed 37 mm.

4. DIFFERENT PROJECTIONS, 7-PARAMETER DATUM TRANSFORM

The main difference of the mappings considered in this section as compared to the previous section is that they include three additional steps:
- transition from geodetic coordinates into Earth-centred Cartesian coordinates;
- applying the 7-parameter datum transform in the Cartesian coordinates;
- projection of the Cartesian coordinates onto the surface of the new reference ellipsoid.

In many GIS applications this is the standard and sometimes the only way to handle datum transforms. While being intuitively understandable and geometrically easy to visualize, this chain is not free of serious problems.

First, the transitions from Earth-centred Cartesian coordinates into geodetic coordinates is computationally expensive. The common textbook formulas (e.g. Strang and Borre, 1997, p. 368) include solving equations with iteration methods. Alternative methods have been proposed to accomplish this task without iterations (see, e.g., Bowring, 1976; Vermeille, 2002; Awange et al., 2010,
Chapter 11 for a thorough overview of various methods) but certain complexity remains in any case.

Second, and more important, the actual parameters in the 7-parameter transforms are calculated from empirical data. Therefore for a given region there may exist many 7-parameter transforms and without additional background information we may not be able to choose the most accurate one.

As an example let us consider mapping from L-EST97 (EPSG:3301) into the Gauss–Krüger 3-degree zone with the central meridian 24° E and Pulkovo 1942 datum (EPSG:2583). We choose this projection because it gives good comparison to the results of the previous section.

We are immediately faced with the problem that there is not one but two transforms in the EPSG parameter set, the domain of validity of which includes Estonia – EPSG:1334 and EPSG:15865. The registry also states that the accuracy of the first transform is 9 m, the accuracy of the second transform is 4.5 m. The maximum difference between these transforms on our test grid is 7.5634 m and the mean square deviation is 6.7600 m. Performing separate computations with each parameter set, we come to the results presented in Tables 3 and 4 and Fig. 3. Software PackageC offered a choice of the transformation and we chose the same transform as we used in direct calculations. PackageA did not offer a choice. It has linked each coordinate system to a fixed datum and uses the transform EPSG:1254 with unknown accuracy for all coordinate systems that use the Pulkovo 1942 datum. Without additional background information we are unable to decide which of the two transformations to prefer. Judging by formal parameters EPSG:15865 would seem a better choice. On the other hand, comparing with reference data from the archives of datasets of the Estonian Land Board – catalogue of triangulation points in the Pulkovo 1942 datum (archive numbers GF-1-10-II-258–GF-1-10-II-283), we see that the datum transform EPSG:1334 produces a mean square deviation of approximately 1.17 m and the datum transform EPSG:15865 respectively 6.83 m. This also demonstrates that the accuracy estimate in the EPSG catalogue for the second transform may be over-optimistic (see also Rüdja, 2004, pp. 204–208).

A closer study of the EPSG transforms reveals some more interesting properties. Comparing the transforms EPSG:1331 (EST92 to ETRS89 (1), the identity transform) and EPSG:1333 (EST92 to WGS 84 (1)), we see that the latter presumes the WGS 84 Doppler realization (cf. McCarthy, 1992, p. 18, Table 3.1). Similarly, comparing the transforms EPSG:1648 (EST97 to ETRS89 (1), the identity transform) and EPSG:1649 (EST97 to WGS 84 (1), the identity transform as well), we see that in this case a more recent realization of the WGS 84 datum is used (G873). This means that transforming coordinates from EST97 into EST92 via WGS 84 would introduce a transformation error of approximately 1 m, whereas it is known that the actual difference between control points of these systems is 4–8 cm (the mean square difference is 4.4 cm, see Lippus, 2004a). Unfortunately, the strategy of conducting all transforms via WGS 84 is used in many widely spread commercial software packages, e.g. PackageA in our comparison.

Table 3. Differences (m) for mappings from L-EST97 into the Pulkovo 1942 Gauss–Krüger 3-degree zone with central meridian 24°, datum transform EPSG:1334

<table>
<thead>
<tr>
<th>Method</th>
<th>Max</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polynomials 3</td>
<td>0.6585</td>
<td>0.2096</td>
</tr>
<tr>
<td>Polynomials 4</td>
<td>0.0078</td>
<td>0.0023</td>
</tr>
<tr>
<td>Polynomials 5</td>
<td>0.0007</td>
<td>0.0003</td>
</tr>
<tr>
<td>PackageA</td>
<td>10.6234</td>
<td>9.6977</td>
</tr>
<tr>
<td>PackageB</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>PackageC</td>
<td>0.0008</td>
<td>0.0007</td>
</tr>
</tbody>
</table>

n/a, not applicable.

Table 4. Differences (m) for mappings from L-EST97 into the Pulkovo 1942 Gauss–Krüger 3-degree zone with central meridian 24°, datum transform EPSG:15865

<table>
<thead>
<tr>
<th>Method</th>
<th>Max</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polynomials 3</td>
<td>0.6585</td>
<td>0.2096</td>
</tr>
<tr>
<td>Polynomials 4</td>
<td>0.0078</td>
<td>0.0023</td>
</tr>
<tr>
<td>Polynomials 5</td>
<td>0.0009</td>
<td>0.0005</td>
</tr>
<tr>
<td>PackageA</td>
<td>5.9762</td>
<td>5.1589</td>
</tr>
<tr>
<td>PackageB</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>PackageC</td>
<td>0.0007</td>
<td>0.0007</td>
</tr>
</tbody>
</table>

n/a, not applicable.

Fig. 3. Distribution of residuals, Gauss–Krüger approximated with polynomials of degree 5. The length of the longest arrow does not exceed 1 mm.
5. IDENTICAL PROJECTIONS, OTHER DATUM TRANSFORMS

Other datum transforms are usually applied on the projection plane of some fixed projection (see, e.g., Dewhurst, 1990; Ordnance Survey U.K., 2005, Chapter 3; Ordnance Survey U.K., 2008, Section 6.3). If we have to work in some other projection, the sequence of steps would be as follows:

- reprojection from the source projection into the fixed projection;
- applying the datum transform;
- reprojection from the fixed projection into the target projection.

To the first and the last step in this sequence we may apply the results of Section 3.

In grid shift methods the area of interest is divided into regular square tiles and we have different mappings on different tiles. We have a table of the values of the shift vectors in the corners of the tiles. The values of the transform in other points are computed by using bilinear interpolation. Bilinear interpolation generally is not conformal, so grid shift transform can be considered nearly conformal if the shift vectors are small.

The software packages that we studied were all able to work with regular grid shift files, but no such files have been published for Estonia. Therefore we cannot continue the comparison of software packages here.

From other methods that can be used in this context we would mention affine mapping on a triangulation and piecewise conformal mapping on convex polygons. The software packages in our comparison are unable to work with these methods.

The triangulation shift datum transform has been used in Finland (see Kuittinen et al., 2006). The difference of this method from the grid shift method is that the grid tiles are triangular (Delaunay triangulation) and affine interpolation is used inside the triangles. Affine interpolation is conformal only inside the triangles where the transformation matrix is orthogonal.

Piecewise conformal mappings with low-degree complex polynomials were studied in Lippus (2004a, 2004b) (see also González-Matesanz and Malpica, 2006). The idea of the method is that if the acceptable accuracy cannot be achieved with a global conformal mapping, we split the area of interest into smaller polygons and solve the conformal mapping problem on each polygon independently. The individual solutions are combined into a global solution by using window functions. The global solution is not conformal in the transition areas from one polygon to another.

6. DISCUSSION AND CONCLUSIONS

We can make three main conclusions that are relevant to map conflation problems.

1. If our data can be transformed with reasonable accuracy, using unprojection, geographic datum transformation, and reprojection, they can be transformed with practically the same accuracy, using complex polynomials of low degree. The degree of the polynomials that are needed depends more on the projections involved than on the datum transformation.

2. In other words: if our data cannot be transformed with reasonable accuracy using complex polynomials of low degree, there exists no 7-parameter transformation that would transform our data with better accuracy.

3. The naïve use of transformations from the EPSG registry, or from the lists of software packages, can introduce certain errors. The risk of such errors can be reduced if the provider of data avoids using potentially ambiguous or little known coordinate systems. For example, if we deliberately label EPSG:3300 data as being EPSG:3301 data, the risk of accidentally introducing the 1 m shift described in Section 4 can be reduced.

The main advantage of complex polynomials is the simplicity of computations.

Remark. Commercial software packages may not fully implement all aspects of the coordinate system as given in the EPSG registry. This concerns especially the order of coordinates. All commercial packages in our comparison assume that the first coordinate in a coordinate pair is Easting and the second is Northing, no matter what is contained in the Coordinate Axes section of the EPSG record.

ACKNOWLEDGEMENTS

This research was partially supported by the Estonian Ministry of Education and Research (targeted financing projects SF0140083s08 and SF0180049s09) and the Estonian Science Foundation (grant 8040). The authors thank the anonymous referees for valuable comments and suggestions.

REFERENCES


Konformsetest georuumilistest teisendustest komplekssette polünoomidega

Jüri Lippus ja Tõnu Oja