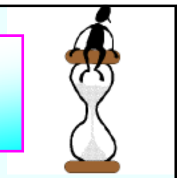
On Choosing the Right Coordinate Transformation Method

Yaron A. Felus[1] and Moshe Felus[2]

[1]The Survey of Israel, Tel-Aviv
Surveying Engineering Department, MI
[2]Halperin Felus Surveying and Photogrammetry

Presentation outline



- **Introduction**
  - The Israeli coordinate based Cadastre project
  - Review of coordinate transformation Procedures
  - Criteria to select the best method
- **Akaike's Information Criterion (AIC)**
- **Case Study in Askelon**
- **More about coordinate transformations and information criteria**
  - The TLS approach
- **Conclusions and further research**

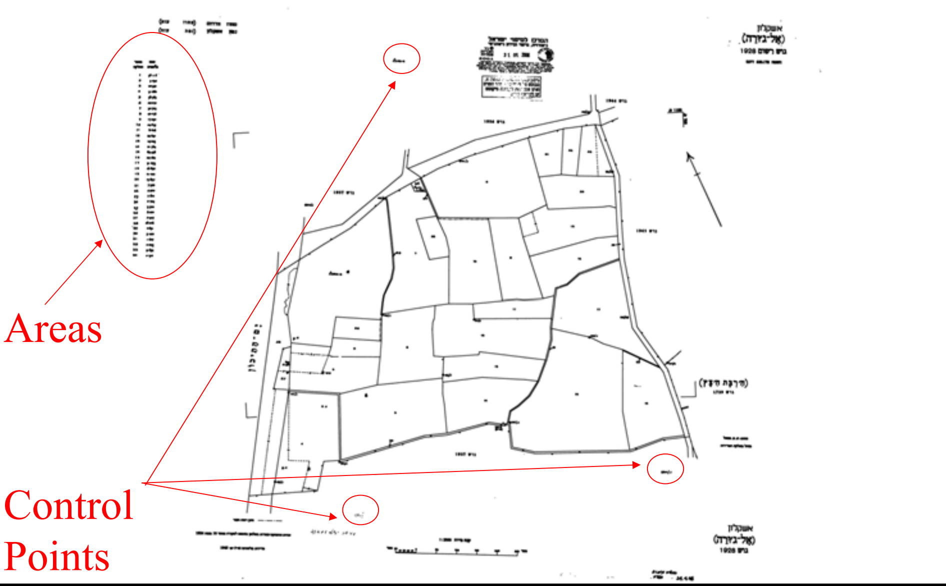
Coordinate Based Cadastre in Israel

In recent years, the Survey of Israel has been evaluating methods and procedure to establish a coordinate based cadastre in Israel.

The coordinate based cadastre is computed from existing maps and datasets and should meet the following requirements:

1. minimize inherent errors in the conversion process,
2. resolve inaccuracies in the original parcel maps,
3. fit the map to the new GPS-based coordinate system with high precision.

Cadastral map



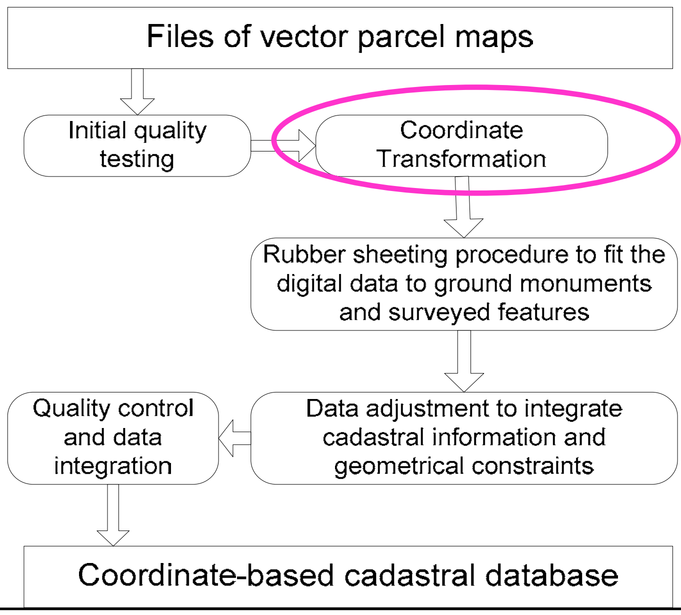
Searching for points using metal detector

Surveying

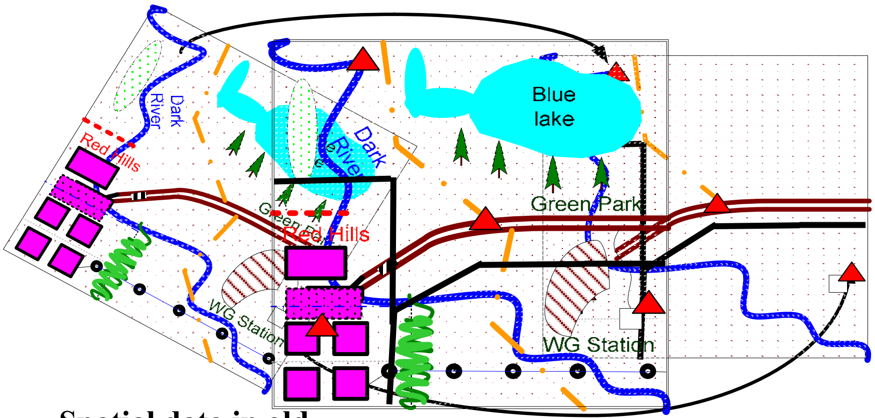
The search for evidence

Surveying of old monuments

Steps in creating a coordinate based cadastre



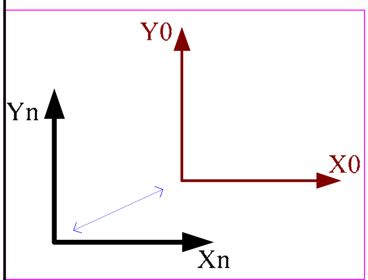
What is a coordinate transformation procedure?



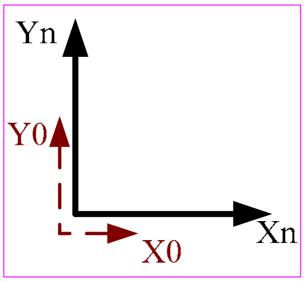
Spatial data in old coordinate system

GPS Control points with uncertainty

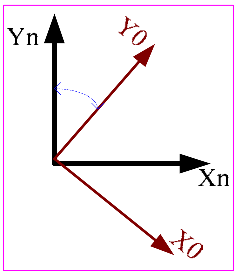
Two-Dimensional Geographic Transformations primitives



Translation



Scaling

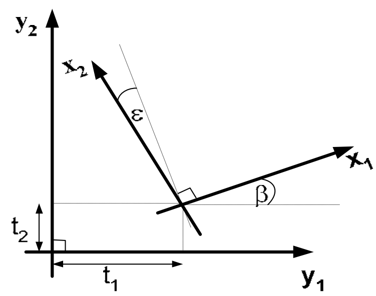


Rotation

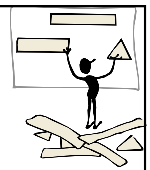
2-D Spatial Data Transformation in the Geosciences

Converting spatial data from a source coordinate system (e.g., image or map coordinate system) to a target coordinate system (e.g., horizontal or object coordinate system).

- Geoscientific applications include:
  - Map conversion,
  - NAD27 to NAD83 geodetic data transformation
  - image orientation



2-D Affine Transformation



- The formulas for an affine transformation:

$$\begin{bmatrix} x_T \\ y_T \end{bmatrix} = \begin{bmatrix} a_A & b_A \\ d_A & e_A \end{bmatrix} \cdot \begin{bmatrix} x_s \\ y_s \end{bmatrix} + \begin{bmatrix} c_A \\ f_A \end{bmatrix}$$

- If n control points are measured, this Equation is reorganized as follows:

$$\underbrace{\begin{bmatrix} x_{T1} \\ y_{T1} \\ \vdots \\ x_{Tn} \\ y_{Tn} \end{bmatrix}}_{y} \approx \underbrace{\begin{bmatrix} x_{s1} & y_{s1} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_{s1} & y_{s1} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{sn} & y_{sn} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_{sn} & y_{sn} & 1 \end{bmatrix}}_{A} \cdot \underbrace{\begin{bmatrix} a_A \\ b_A \\ c_A \\ d_A \\ e_A \\ f_A \end{bmatrix}}_{\xi}$$

Isogonal Affine Transformation or Conformal/Similarity Transformation

- Isogonal: having equal angles
- Impose additional condition of equal scale ($S = C_x = C_y$) yielding <u>4 parameters</u>: S, α, DX_0, DY_0

$$\begin{bmatrix} X_T \\ Y_T \end{bmatrix} = s \cdot \begin{bmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{bmatrix} \cdot \begin{bmatrix} X_o \\ Y_o \end{bmatrix} + \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \cdot \begin{bmatrix} X_o \\ Y_o \end{bmatrix} + \begin{bmatrix} c \\ d \end{bmatrix}$$

Projective or Polynomial Transformation

- Instead of 4 or 6 parameters we have many parameters at least 8 (8 would be the Bi-linear or projective transformation)
- With more parameters we need more known points to solve the equations
- N-equations and N unknowns.

$$Xn = a_0 + a_1 X_0 + a_2 Y_0 + a_3 Y_0 X_0 + a_4 X_0{}^2 + a_5 Y_0{}^2 + \ldots$$

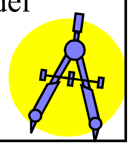
$$Yn = b_0 + b_1 X_0 + b_2 Y_0 + b_3 Y_0 X_0 + b_4 X_0{}^2 + b_5 Y_0{}^2 + \ldots$$

What transformation model to use?

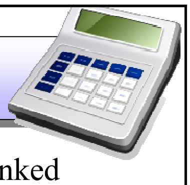
- **Translation, Similarity, Affine, Projective, or Polynomial.?**
- The more parameters we have, the smaller the RMSE. – **Good**
- More parameters mean more distortion are introduced to the system – **Bad.**
- Chen and Hill(2005) proposed the following criteria: invertability, precision - as measured by the Root Mean Squared (RMS) error of the residuals, the maximum residual and the 95% of the available residuals, uniqueness, conformality, and extensibility.
- he analysis of Chen and Hill, 2005 considered many aspects of the transformation and concluded that a polynomial model is the best choice; nonetheless, a combined factor that incorporates all these criteria was not suggested in that research.

The Akaike's Information Criterion (AIC)

- Akaike's information criterion, developed by Hirotsugu Akaike under the name of "an information criterion" (AIC) in 1971 and proposed in Akaike (1974), is a measure of the goodness of fit of an estimated statistical model.
- It is a relative measure of the information lost when a given model is used to describe reality. The tradeoff between the accuracy and complexity of the model.
- The Akaike's Information Criterion (AIC) replaces the previous methods that relies on hypotheses testing, to select the model that is optimal (Kullback–Leibler theorem).



AIC – The Math



- Given a data set, several competing models may be ranked according to their AIC, with the one having the lowest AIC being the best.

$$AICc = n \cdot \log(WRSS) + 2 \cdot k \cdot (n/(n-k-1))$$

- n – number of observation
- k - number of parameters
- WRSS – Weighted sum of squared residuals.

$$WSSR := \tilde{\mathbf{e}}^{\mathbf{T}} \cdot \bar{\mathbf{P}} \cdot \tilde{\mathbf{e}}^{\mathbf{T}}$$

$$\tilde{\mathbf{e}} = \mathbf{y} - \mathbf{A}\hat{\xi}$$

$$\hat{\xi} = (\mathbf{A}^{\mathbf{T}}\mathbf{PA})^{-1}\mathbf{A}^{\mathbf{T}}\mathbf{Py}$$

Case study of choosing the best model

Translation, Similarity, Affine, Projective Polynomial?

An area with 19 control points in two systems (n=38) was selected

	Translation	Similarity	Affine	Projective
Number of Parameters (k)	2	4	6	8
WRSS	0.9524	0.544	0.491	0.4519
AICc	-48.103	-51.891	-45.176	-34.323

The similarity transformation presents an optimal balance of information content and model complexity.

AICc used for data selection

- The AIC criterion can be employed as a criterion for point selection and the elimination of bad information.
- Traditionally, in Geodesy we employ all the given data with the exception of outliers.
- The definition of an outlier is subjective (selection of statistical significance level 99%-> 80%) as used by Baarda's data snooping.
- Two approached are investigated to obtain "optimal" balance of information content and model accuracy.
  - Eliminates points with low significant
  - Eliminates points with high residuals

How to detect an outliers:

According to Chebyshev theorem almost all the residuals in a data set are going to be in the interval

$$(\bar{Z} - 3 \times SD, \quad \bar{Z} + 3 \times SD)$$

where

$\bar{Z}$ is the mean

SD is the standard deviation of the sample.

Therefore, the observations with residual outside this interval will be considered as an outliers.

This process does not consider the different weights of the residuals.

Outlier detection process

Detecting an outlier using Baarda's data snooping procedure:

- Compute the residual vector $\tilde{v} = f - B\hat{\Delta}$
- Compute the residual vector cofactor matrix

$$Q_{\tilde{v}} = W^{-1} - B \cdot (B^T W B)^{-1} \cdot B^T$$

- Compute the posteriori standard deviation (reference variance)

$$\sigma_0 = \sqrt{\frac{v^t \cdot W \cdot v}{r}}$$

- Compute the standardized residual and test if it is an outlier by:

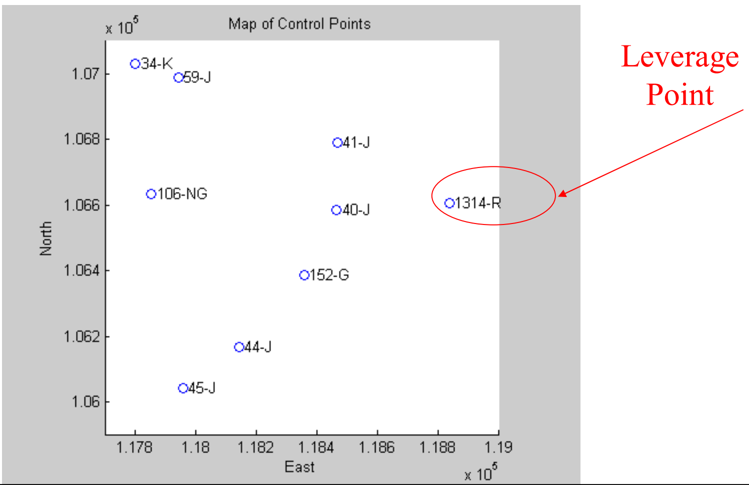
$$\bar{v}_i = \frac{v_i}{\sigma_0 \cdot \sqrt{q_{ii}}} > 2.8$$

- This is a rejection criterion which corresponds to (1-α)=0.95

Reference: Wolf & Ghilani pp 404-406

Leverage points and insignificant points

- The diagonal elements of the Hat matrix ($\mathbf{H} := \mathbf{A}(\mathbf{A}^T\mathbf{P}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{P}$) identify insignificant points. Insignificant points may not be used if the AIC suggests so.



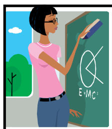
AICc used for point selection

Optimal number of points removing insignificant points

Experiment	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
No of Points	9	8	7	6	5	4	3	2
WRSS	0.544	0.510	0.065	0.050	0.029	0.022	0.022	0
AICc	-51.89	-43.48	-62.56	-51.83	-42.36	-25.65	14.43	∞
Point removed	-	152-G	40-J	16NG	41-J	45-J	59-J	44-J

Optimal number of control points for Similarity transformation by eliminating the point with the largest discrepancy.

Experiment	1	2	3	4	5	6	7	8
No. of points	9	8	7	6	5	4	3	2
WRSS	0.54	0.07	0.05	0.04	0.02	0.01	0.007	0
AICc	-51.89	-74.68	-65.48	-54.59	-43.89	-29.60	7.783	∞
Point removed	-	40-J	44-J	41-J	34-K	59-J	152-J	16NG



Other Information Criteria

- The Mallows statistic *Cp* (Mallows, 1973), may be used as well. The Mallows statistic is given by:

$$Cp = (\hat{\sigma}^2)^{-1} \cdot (WRSS) \;-\; n + 2 \cdot k$$

- where is a properly chosen estimate of the posteriori reference variance, *n* is the number of observations, and *k* is the number of parameters. However, the Mallows statistic *Cp* was criticized as being subjective to the choice of the posteriori reference variance
- The Bayes Information Criterion (Schwarz, 1978).

$$BIC := n \cdot \log(WSSR) + \log(n) \cdot k$$

- The BIC is an increasing function of WRSS and an increasing function of k.

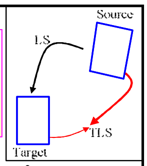
Concluding remarks and further research

Key results:

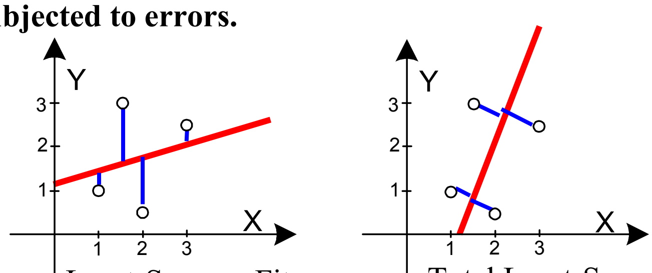
- Use of the AIC for transformation model selection
- Use of the AIC for outlier detection still under investigation
- Use of BIC still under investigation

Comments/Questions?

Coordinate Transformation with the TLS approach



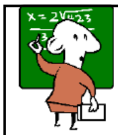
The Total Least-Squares approach is concerned with estimating parameters using the error in all variables model namely both the observation vector y and the data matrix A are subjected to errors.



Least-Squares Fit   Total Least-Squares Fit

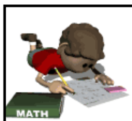
TLS algorithms were used for coordinate transformations, variogram estimation and trend estimation.

Significant accuracy improvement was achieved.



Total Least Squares(TLS) Problem

- Given an overdetermined set of linear equations $\mathbf{y} \approx \mathbf{A}\xi$
- where
  - **y** is the observation vector
  - **A** is a positive defined data matrix,
  - ξ is the vector of unknown parameters.
- The Total Least Squares problem is concerned with estimating ξ, providing that the number of observations (n) is larger than the number of parameters (m) to be estimated, and given that **both the observation vector y and the data matrix A are subjected to errors**



Total Least Squares(TLS) Problem

Total Least-Squares (TLS) is a method to estimate parameters in linear models that include random errors in all their variables.

$$(A - E_A) \cdot \xi - (y - e) = 0, \qquad \textbf{(1)}$$

$$E\{[E_A, e]\} = 0, \quad D\{vec[E_A, e]\} = \Sigma_0 \otimes I_n, \quad C\{E_A, e\} = 0.$$

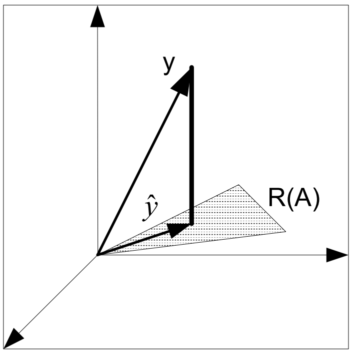
Where:

- E_A is the random error matrix associated with the data matrix
- e is a random error vector associated with the observation vector.
- The “vec” operator stacks one column of a matrix under the other, moving from left to right.
- $\Sigma_0 = \sigma_0^2 \cdot I_{m+1}$ is a (m+1)×(m+1) matrix with unknown variance component σ_0^2 and given identity matrix I_{m+1}.

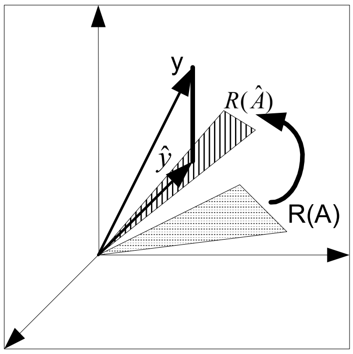
Total Least Squares(TLS) Solution

The TLS principle is based on minimizing the following objective function:

$$e^T e + (vec\,E_A)^T (vec\,E_A) = \min(\xi)$$



Least-Squares approach

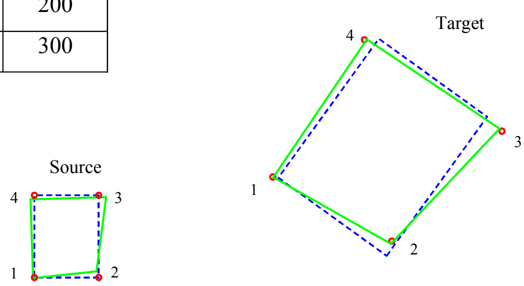


Total Least-Squares approach

Case study: Map georeferencing

Two digital maps of the same area need to be merged, using a set of control points.

Point No.	Source map x_T	Source map y_T	Target map x_S	Target map y_S
1	30	40	290	150
2	100	40	420	80
3	100	130	540	200
4	30	130	390	300



Results Map georeferencing

Affine transformation parameters, the LS vs. TLS

Parameter	a_A	b_A	c_A	d_A	e_A	f_A
LS	2.0000	1.2222	176.1111	-1.2142	1.5000	133.9285
TLS	2.0126	1.2185	175.6018	-1.2332	1.5055	134.6924

Similarity transformation parameters, the LS vs. TLS

Parameter	a	b	c	d
LS	1.6884	1.2192	196.6153	118.2307
TLS	1.7150	1.2384	193.2582	117.2196

Sum of squared errors of the different adjustment methods

	Affine transformation TLS method	Affine transformation LS method	Similarity transformation TLS method	Similarity transformation LS method
Sum of Squared Errors	57.3591	325.00	201.3557	1088.5