Homogenisierung von Massendaten im Kontext von Geodaten-Infrastrukturen

Stefan Kampshoff und Wilhelm Benning

Zusammenfassung

Der Beitrag beschreibt zunächst das geodätische Verfahren der Homogenisierung, das seinen Ursprung in der Datenaufbereitung im Zuge der Digitalisierung analoger Karten hat. Der Zusammenhang zu den aus der Geoinformatik stammenden Ansätzen der Integration und Verschmelzung heterogener räumlicher Datenquellen wird aufgezeigt. Es folgt eine Diskussion des Begriffs der geometrischen Heterogenität als Folge der Fragmentierung von Geodatenbeständen. Zuletzt wird gezeigt, dass das Verfahren der Homogenisierung die Integration geometrisch heterogener Datenquellen ermöglicht und dadurch einen Beitrag zur Herstellung geometrischer Interoperabilität in Geodaten-Infrastrukturen leisten kann.

Der zweite Teil des Beitrags stellt das Modell der Homogenisierung und seine Umsetzung in dem Programmsystem KATHOM vor. Die Lösung eines Normalgleichungssystems, dessen Dimension durch die Anzahl der nicht-inzidenten Koordinaten der zu integrierenden Geometrien (Neupunkte) determiniert ist, bildet den Flaschenhals der Homogenisierung von Massendaten. Es folgt eine Untersuchung verschiedener Ansätze zur Überwindung des Flaschenhalses und deren Vergleich anhand realer Datensätze. Praktische Beispiele zeigen, dass ein Fill-In minimierendes Sortierverfahren, das einen Multilevel Nested Dissection Algorithmus (MLND) verwendet, zu einer Verbesserung der Performanz der Homogenisierung von Massendaten bis zum Faktor 80 gegenüber den bisher in KATHOM verwendeten Ansätzen führt.

Summary

This paper describes the geodetical procedure of homogenisation, that has its origin in the processing of digitised analogous maps. The link to the geoinformation-science problem of integration and conflation of heterogeneous data sources is established. We discuss geometrical heterogeneity as a consequence of the fragmentation of spatial datasets. It is shown, that the procedure of homogenisation enables the integration of spatially heterogeneous data sources, and may thereby contribute to geometrical interoperability in spatial data infrastructures (SDI).

The second part of this paper presents the functional model of homogenisation and its implementation in the program KATHOM. The solution of a system of normal-equations is identified as the bottleneck of the problem. The number of unknown parameters is determined by the number of non-incident coordinates of the geoobjects that have to be integrated. Multiple approaches to overcome this bottleneck are investigated and compared using real data sets. Practical examples show, that a fill-in minimising ordering procedure, that uses a multi-level nested dissection approach, improves the performance of the homogenisation of large data sets up to the factor 80 compared to the approaches that have been used before in KATHOM.

1 Einleitung

Der Begriff der Homogenisierung umschreibt die Verkettung nichtlinearer topologischer Transformationen (auch: Gummituch oder *rubber-sheet* Transformation), mittels derer heterogene Mengen von Punkten, Linien und Flächen[1] aus unterschiedlichen Quellen nachbarschaftstreu und unter Beibehaltung datenimmanenter geometrischer Beziehungen in ein gemeinsames homogenes Zielsystem transformiert werden.

Die Zusammenführung heterogener Objekte aus unterschiedlichen Datenquellen zu einem konsistenten Ganzen bezeichnet man gemeinhin als (Daten-)Integration (Kampshoff und Benning 2002). Die Homogenisierung ist ein Sonderfall dieser allgemeinen Integration; sie ist beschränkt auf die Geometrien raumbezogener Objekte. Die Homogenisierung wird daher – in Abgrenzung zu semantischer, schematischer und syntaktischer Integration – auch als *geometrische Integration* bezeichnet (Walter 1997, Hake et al. 2002).

Die Homogenisierung stellt somit ein Verfahren zur geometrischen Integration dar, das der Herstellung *geometrischer Interoperabilität* zwischen räumlichen Datenquellen dient.

Das klassische Anwendungsgebiet der Homogenisierung liegt in der Datenaufbereitung bei der Erfassung großmaßstäbiger digitaler Datenbestände aus analogem Kartenmaterial (Benning und Scholz 1990, Gielsdorf et al. 2003, Hake et al. 2002, Hampp 1992, Morgenstern et al. 1988, Rose 1988, Wiens 1984). Das Ziel besteht dabei in der Einpassung und verketteten Transformation digitalisierter Karten (Startsysteme) in das übergeordnete Landessystem (Zielsystem). Dabei entstehen Startsysteme von der Größe etwa eines Blattschnitts oder einer Inselkarte.

In der Fortführungspraxis digitaler Geodatenbestände ermöglicht die Homogenisierung die permanente nachbarschaftstreue Anpassung bestehender räumlicher Objekte an verbesserte Anschlusskoordinaten (Benning 1996, Liesen 1998). Die bei der Fortführungshomogenisierung auftretenden Datenmengen sind durch den Umfang der Transaktionen begrenzt auf einige hundert, in seltenen Fällen auch einige tausend Objekte.

1 Für den dreidimensionalen Fall ist die Definition um Volumen zu erweitern, siehe z. B. Müller 1999.

Die Umstellung des Lagebezugssytems in großen Geobasisdatenbeständen (Netztransformation bzw. Lagestatuswechsel) stellt besondere Anforderungen an die Skalierbarkeit der Homogenisierung (Kampshoff und Benning 2002).

Dem Wechsel des Lagebezugssystems in Primärdatenbeständen folgt eine Aktualisierung der Sekundärdatenbestände auf der Seite der Nutzer von Geobasisdaten (z. B. Energieversorger, Kommunen). Das betrifft insbesondere solche Nutzer amtlicher Geobasisdaten, die über eine Dokumentation ihrer Vermögenswerte und Betriebsmittel auf der Grundlage amtlicher Geobasisdaten verfügen (Gielsdorf et al. 2004, Liesen 1998, Scheu et al. 2000, Stockwald 2000). Die Änderungen an den Geobasisdaten zerstören die (impliziten) geometrischen Relationen zwischen Basiskarte und Benutzerobjekten (Assoziativitätsproblem) (Wan und Williamson 1994a, Wan und Williamson 1994b). Die Aufgabe der Homogenisierung besteht darin, die Konsistenz zwischen Geobasisdaten und Benutzerdaten wiederherzustellen.

Die Herausforderung für die Homogenisierung bei großflächigen Netzänderungen liegt in der Bewältigung der anfallenden Massendaten, die leicht zu einer Größenordnung von einigen Millionen Punkten führen können.

2 Geometrische Interoperabilität in Geodaten-Infrastrukturen

Innerhalb von Geodaten-Infrastrukturen ermöglichen Dienste und Formate wie WFS (Web Feature Service) und GML (Geography Markup Language) den Zugriff auf syntaktisch heterogene Datenquellen. Die gemeinsame Nutzung integrierter Datenbestände erfordert neben der syntaktischen Interoperabilität die Überwindung geometrischer Heterogenität (Gröger und Kolbe 2003). Es besteht daher Bedarf nach einem Dienst für die geometrische Integration, dessen Implementierung auf der Basis des geodätischen Verfahrens der Homogenisierung erfolgen kann.

Im Unterschied zu dem klassischen Einsatzgebiet der Homogenisierung handelt es sich dabei um ein online-Verfahren, bei dem die Entscheidung über zu integrierende Datenbestände und -modelle ad hoc getroffen werden kann (Bernard et al. 2003). Der Dienst erfordert über eine hohe Flexibilität und Konfigurierbarkeit für Daten aus unterschiedlichen Modellen hinaus ein hohes Maß an Performanz und Skalierbarkeit, um als server-seitige Lösung in einem online-Verfahren in Betracht zu kommen.

Riedemann und Timm formulieren in diesem Zusammenhang das ehrgeizige Ziel einer automatischen Echtzeit-Integration *(automated just-in-time integration)* auf der Basis einer System-Architektur verteilter Daten und Dienste (Riedemann und Timm 2003). Zur Integration soll in einem ersten Schritt die Definition eines gemeinsamen Modells *(common model)* erfolgen, auf das sämtliche Datenbestände abgebildet werden, vergleichbar mit der Definition eines globalen Schemas in föderierten DBMS (Conrad 1997, Kampshoff und Benning 2002). In einem zweiten Schritt findet eine Korrespondenzanalyse der Datensätze, sowohl auf der Modellebene (Schematransformation) als auch auf der Instanzebene, statt.

Die Suche nach korrespondierenden Instanzen führt, soweit keine anderen identifizierenden Merkmale vorhanden sind, zu einem geometrischen Zuordnungsproblem. Im Anschluss an die Zuordnung hat eine nachbarschaftstreue Anpassung der integrierten Datenbestände zu erfolgen (Laurini 1994).

Die klassischen Homogenisierungsverfahren auf dem Gebiet der Datenaufbereitung umfassen sowohl geometrische Zuordnungsalgorithmen als auch Techniken zur nachbarschaftstreuen Transformation und Anpassung. Die Verwendung der bestehenden Ansätze als *Processing-Service* einer Dienste-Architektur für die Datenintegration innerhalb von Geodaten-Infrastrukturen liegt daher nahe (OGC 2003). Die genannten Anforderungen zur Flexibilität und Konfigurierbarkeit könnten ein Hindernis bei der Wiederverwendung bestehender Homogenisierungslösungen darstellen, da diese oftmals eng an die Strukturen proprietärer GIS gekoppelt sind. Eine Standardisierung von Dienste-Schnittstellen in diesem Bereich ist hilfreich, denn obgleich schon in Kottman (2000) die Notwendigkeit von Diensten zur Geodaten-Integration *(data fusion services)* aufgezeigt wurde, ist dieser Bereich in der internationalen Standardisierung (ISO, OGC etc.) bisher wenig fortgeschritten.

Ein Dienst zur Echtzeit-Homogenisierung verteilter Geodatenbestände war bis vor wenigen Jahren allein aufgrund der benötigten Rechenzeiten undenkbar. Aufgrund der allgemeinen Leistungssteigerung im Bereich der Hardware, vor allem aber aufgrund der rasanten Weiterentwicklung der relevanten Verfahren der numerischen Mathematik, die in Abschnitt 4 dieses Beitrags vorgestellt wird, ist sie jedoch in greifbare Nähe gerückt.

2.1 Abstraktionsebenen der geometrischen Heterogenität

Semantische, schematische und syntaktische Heterogenität, wie sie in Bishr definiert werden, sind struktureller Art und betreffen die Ebene des konzeptuellen Modells (Gröger und Kolbe 2003, Bishr 1998). Die Ausprägungen geometrischer Heterogenität hingegen können unterteilt werden in global-systematische, lokal-systematische und zufällige Effekte.

Die geometrische Heterogenität mit global-systematischer Natur betrifft sämtliche Instanzen einer Datenquelle in gleicher systematischer Weise. Sie kann somit der Modellebene zugeordnet werden. Ein Beispiel ist die systematische Abweichung, die aus der fehlerhaften Einpassung gescannter Karten in das Landessystem resultiert. Räumliche Objekte, die aus dieser Datenquelle gebildet werden, sind in gleicher Weise systematisch verfälscht.

Neben den global-systematischen Abweichungen existieren lokal-systematische und zufällige Abweichungen auf der Ebene einzelner Instanzen bzw. Geometrien. Ein Beispiel ist die lokal begrenzte Verzerrung eines Datenbestandes, die beispielsweise durch den Anschluss einer lokalen Vermessung an einen fehlerhaft koordinierten Aufnahmepunkt oder durch einen lokalen Papierverzug bei der Digitalisierung analoger Karten verursacht sein kann.

Im Ergebnis ist festzuhalten, dass geometrische Heterogenität räumlicher Objekte sowohl auf der lokalen Ebene einzelner Instanzen oder Cliquen benachbarter Instanzen als auch auf der globalen Modellebene auftritt.

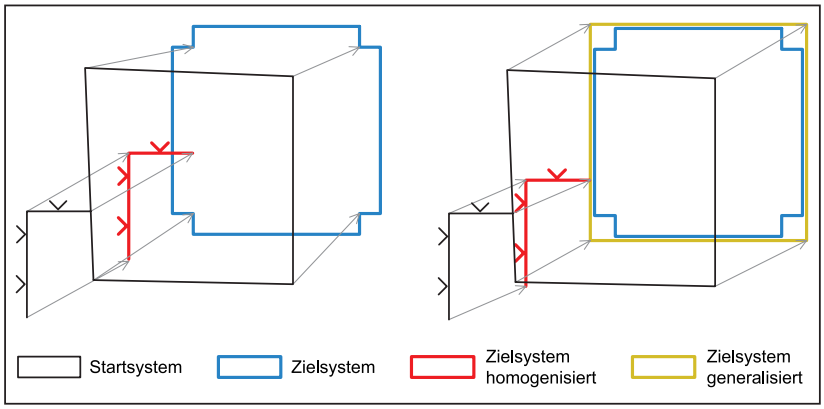

Abb. 1: Geometrische Integration bei heterogenem Detaillierungsgrad ohne (links) und mit (rechts) Generalisierung

2.2 Ausprägung geometrischer Heterogenität

Die Heterogenität der Geometrien aus unterschiedlichen Quellen zeigt sich auf der Ebene einzelner Punkte darin, dass die Koordinaten homologer Punkte aus unterschiedlichen Datensätzen voneinander abweichen. Des Weiteren können Inkonsistenzen hinsichtlich der geometrischen und topologischen Relationen der Geometrien unterschiedlicher Systeme auftreten. Dabei unterscheidet man zwischen formalen Inkonsistenzen, die durch eine Überprüfung der inneren logischen Struktur der Daten aufgedeckt werden können (Gröger 2000), und Abbildungsfehlern, bei denen zwar eine formal korrekte Instantiierung des Datenmodells vorliegt, die jedoch den erfassten Zustand der realen Welt nicht korrekt widergibt.

Da es sich bei der Homogenisierung ausschließlich um ein geometrisches Verfahren handelt, spielen Abweichungen in der Granularität des Objektmodells und der Objektbildung eine untergeordnete Rolle, solange die Geometrien der Datenbestände strukturelle Ähnlichkeiten aufweisen. Maßstabsunterschiede, die zu unterschiedlichen Detaillierungsgraden der räumlichen Objekte führen, sind hingegen problematisch für die geometrische Integration. Bei kleineren Unterschieden kann oftmals eine Anpassung mit einfachen Generalisierungsalgorithmen erfolgen, sodass eine Integration der Objekte auf gleicher Detaillierungsstufe stattfinden kann. In dem Beispiel aus Abb. 1 ist eine Vereinfachung des Gebäudeumrings im Zielsystem erforderlich, um eine konsistente Anpassung der Benutzerdaten (hier: Topographie/Zaun) an die Objekte des Zielsystems zur gewährleisten. Zunächst existieren zu den Gebäudeeckpunkten des Startsystems keine homologen Objekte im Zielsystem, so dass die Topographie fehlerhaft homogenisiert wird (linke Seite der Abb. 1). Im vorliegenden Fall kann der Gebäudeumring im Zielsystem durch Generalisierung so angepasst werden, dass die Topographie korrekt homogenisiert wird (rechte Seite der Abb. 1).

Bei starken Maßstabsdifferenzen, die beispielsweise zu einem Wechsel des Geometrietyps homologer Objekte führen können, kommt eine Anpassung mittels Homogenisierung nicht in Betracht, da sich keine topologische Transformation angeben lässt, die beispielsweise Flächen in Punkte überführt.

2.3 Ursache geometrischer Heterogenität: Fragmentierung von Geodaten

Die Ursache geometrischer Heterogenität liegt in der organisatorisch und physisch getrennten Erfassung, Fortführung und Haltung von Geodaten, die zu einer Zersplitterung von Information geführt hat. Grundsätzlich unterscheidet man zwei Ausprägungen der Zersplitterung, die *Layer-Fragmentierung* und die *zonale Fragmentierung* (Laurini 1998, Gröger und Kolbe 2003):

- Layer-Fragmentierung tritt zwischen Objekten unterschiedlicher Klassen aus Datenbeständen mit kongruenter oder stark überlappender räumlicher Ausdehnung auf, beispielsweise bei der Kombination von Gas- und Wasserdaten, die auf der Grundlage unterschiedlicher Geobasisdaten referenziert wurden.
  Nahezu jede Aktualisierung der Geometrie von Geobasisdaten führt auf der Seite der Datennutzer zu einer Layer-Fragmentierung. Integrierte Nutzerdatenbestände enthalten neben den Geobasisdaten weitere Layer von Objekten. So zieht die Aktualisierung der Geobasisdaten in Netzinformationssystemen eine Layer-Fragmentierung zwischen den aktualisierten Geobasisobjekten und den Objekten der Netzsparten nach sich (Scheu et al. 2000, Stockwald 2000).
  Die Layer-Fragmentierung wird auch als *thematische Partitionierung* und im Kontext relationaler DBMS als *vertikale Fragmentierung* bezeichnet (Worboys und Duckham 2004, S. 276).
- Als zonale Fragmentierung bezeichnet man die Aufteilung der Objekte einer Klasse in Datenbestände, die räumlich benachbarte weitestgehend disjunkte Gebiete umfassen. Ein klassisches Beispiel zonaler Fragmentierung stellen die geometrischen Abweichungen benachbarter Inselkarten dar, die beim Aufbau digitaler Geodatenbestände zusammengeschlossen werden müssen.

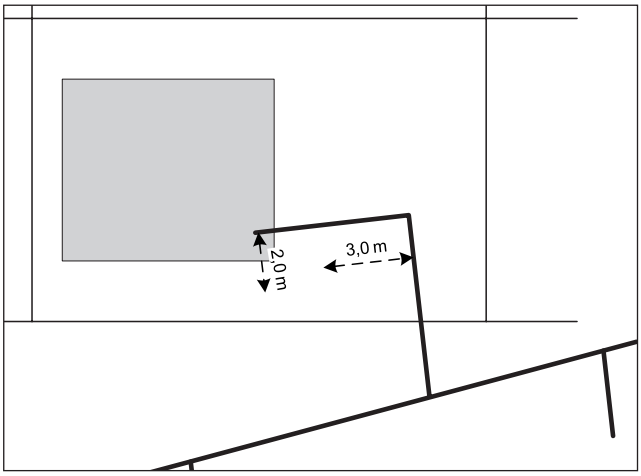

Abb. 2: Layer-Fragmentierung zwischen Geobasisdaten und Netzdaten

Eine ähnliche Aufgabe stellt sich bei der Kombination von Geobasisdaten verschiedener administrativer Gebietseinheiten (Kreise, Bundesländer oder Staaten) (Illert 1995). Sie können ebenfalls zonal fragmentierte Informationen enthalten, deren Widersprüche durch geometrische Integration beseitigt werden müssen.
Die zonale Fragmentierung nennt man auch *räumliche Partitionierung* und im Kontext relationaler DBMS *horizontale Fragmentierung*.

2.4 Homogenisierung: Beseitigung von Fragmentierung als Transformationsproblem

Der Zusammenhang zwischen den Geometrien fragmentierter Datenbestände ergibt sich aus den Koordinaten homologer Punkte und aus Datenbestand übergreifenden geometrischen Relationen.

In der Homogenisierung werden die Geometrien der einzelnen Datenbestände jeweils einem Koordinatensystem zugeordnet. Dabei können multiple Startsysteme auftreten, denen die Geometrien fragmentierter Datenquellen entstammen. Daneben existiert in der Regel ein ausgezeichnetes Zielsystem. Für Anschlusspunkte (auch: Passpunkte) liegen Koordinaten zugleich im Start- und Zielsystem vor. Punkte, deren Koordinaten in mehreren benachbarten Startsystemen vorliegen, werden als Rand- oder Verknüpfungspunkte bezeichnet. Die dritte Gruppe von Punkten bilden die Neupunkte, deren Koordinaten in nur einem der Startsysteme vorliegen. Darüber hinaus können sich aus geometrischen Beziehungen implizite Zusammenhänge zwischen den Systemen ergeben, beispielsweise in Form von Geradheits- oder Abstandsrelationen, an denen Punkte mit Koordinaten aus unterschiedlichen Systemen beteiligt sind.

Die beschriebenen Zusammenhänge lassen sich in ein objektorientiertes Datenmodell umsetzen, das in Abb. 3 als statische UML-Struktur dargestellt ist. Jeder Datenbestand-Referenz ist ein Startsystem zugeordnet. Dabei können multiple Datenbestände einem Startsystem angehören, wenn sie geometrisch zueinander konsistent sind (z. B. geometrisch konsistente Sparten eines Netzinformationssystems). Ein Punkt kann aus einer Startkoordinate (Neupunkt), multiplen Startkoordinaten (Verknüpfungspunkt) oder einer Zielkoordinate und beliebig vielen Startkoordinaten (Anschlusspunkt) bestehen. Es gibt einige Nebenbedingungen, die in der Darstellung nicht enthalten sind. Beispielsweise sind die Instanzen der Koordinaten-Klassen innerhalb eines Systems durch ihre Koordinatenwerte eindeutig identifiziert und Punkte mit multiplen Startkoordinaten aus einem Startsystem sind unzulässig.

Das zentrale Ziel der Homogenisierung ist die Integration fragmentierter räumlicher Datenbestände durch Beseitigung der geometrischen Heterogenität zwischen den Geometrien unterschiedlicher Systeme. Übertragen auf das Datenmodell der Abb. 3, liegt die Aufgabe der Homogenisierung darin, Zielkoordinaten für sämtliche Instanzen der Klasse Punkt zu ermitteln.

Dazu wird für jedes Startsystem S_i eine topologische Transformation $T^Z_{S_i} : S_i \rightarrow Z$ von dem Startsystem in das Zielsystem ermittelt, wobei die Beziehungen benachbarter Startsysteme berücksichtigt werden. Das kann durch eine gemeinsame Bestimmung aller Transformationen $T^Z_{S_1}, \ldots, T^Z_{S_n}$ in einem Ausgleichungsprozess geschehen, wobei die Verkettung der Transformationen durch Verknüpfungspunkte und geometrische Bedingungen erfolgt.

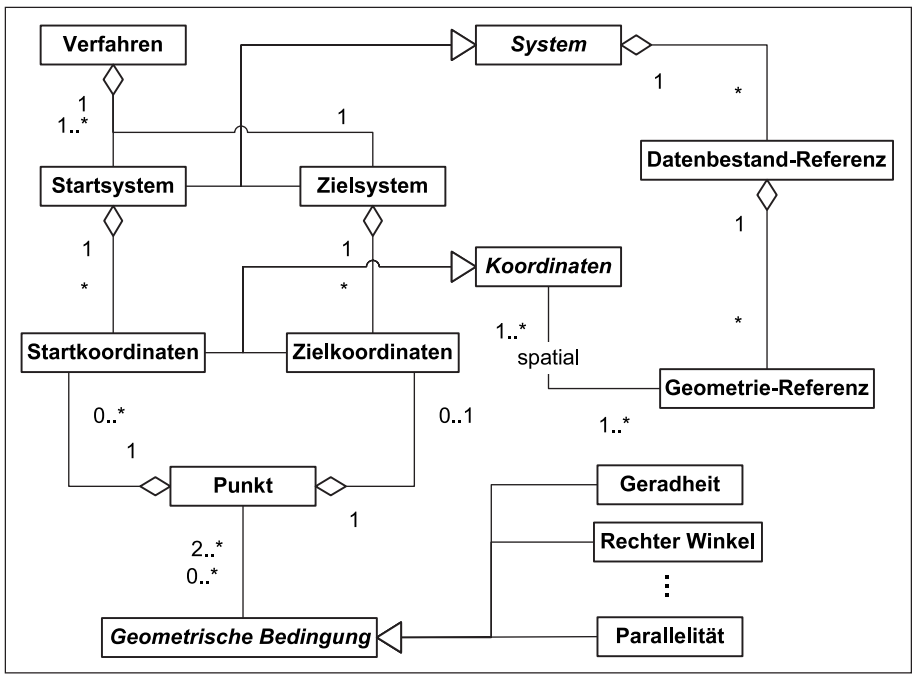

Abb. 3: Statisches Modell der Homogenisierung

3 Das Ausgleichungs-Modell der Homogenisierung

Das funktionale Modell der Homogenisierung im Programmsystem KATHOM ist in Benning (1995), Benning (1996), Hettwer und Benning (2000), Hettwer (2003) beschrieben. An dieser Stelle wird es daher ausschließlich in den Grundzügen dargestellt.

3.1 Die Beobachtungsgleichungen der Homogenisierung

Die Homogenisierung wird als Parameterschätzung im Gauß-Markoff-Modell formuliert (Koch 1997), dessen u unbekannte feste Parameter $\boldsymbol{\beta}$ sich zusammensetzen aus

- $\boldsymbol{\beta}_z$, den k unbekannten Zielsystem-Koordinaten der Neupunkte, und
- $\boldsymbol{\beta}_t$, den l unbekannten Transformationsparametern einer oder mehrerer (o. B. d. A.) 5-Parameter Transformationen der Startsysteme in das Zielsystem.

Die n Beobachtungen des Gauß-Markoff-Modells der Homogenisierung entstammen drei verschiedenen Gruppen, die hier zusammengefasst sind in den Vektoren

- $\boldsymbol{y}_s$, der m Koordinaten der Punkte in einem oder mehreren Startsystemen (Hettwer 2003, S. 40),
- $\boldsymbol{y}_b$, der o Beobachtungen zum Zwecke der Realisierung geometrischer Bedingungen (Hettwer 2003, S. 45), und
- $\boldsymbol{y}_n$, der q fingierten Beobachtungen zwischen benachbarten Punkten nach dem hybriden Ansatz (Benning 1995; Hettwer 2003, S. 55), die eine nachbarschaftstreue Transformation der Punkte gewährleisten.

Die zugehörigen Beobachtungsgleichungen $y_i + e_i = h_i(\boldsymbol{\beta})$ sind zu großen Teilen nicht-linear, sodass für den Übergang in das lineare Gauß-Markoff-Modell eine Linearisierung der Beobachtungsgleichungen mit geeigneten Näherungswerten $\boldsymbol{\beta}^{(0)}$ der unbekannten Parameter erfolgt.

In Blockmatrizen-Schreibweise erhält man das lineare Gauß-Markoff-Modell der Homogenisierung bei Annahme unabhängiger Beobachtungen zu

$$E\left(\begin{vmatrix}\boldsymbol{y}_s\\ \boldsymbol{y}_b\\ \boldsymbol{y}_n\end{vmatrix}\right) = \begin{vmatrix}\boldsymbol{X}_{11} & \boldsymbol{X}_{12}\\ \boldsymbol{X}_{21} & \boldsymbol{0}\\ \boldsymbol{X}_{31} & \boldsymbol{0}\end{vmatrix} \cdot \begin{vmatrix}\boldsymbol{\beta}_z\\ \boldsymbol{\beta}_t\end{vmatrix} \quad \text{mit} \tag{1}$$

$$D\left(\begin{vmatrix}\boldsymbol{y}_s\\ \boldsymbol{y}_b\\ \boldsymbol{y}_n\end{vmatrix}\right) = \sigma^2 diag(p_1,\ldots,p_m,p_{m+1},\ldots,p_{m+o}, p_{m+o+1},\ldots,p_{m+o+q})^{-1}. \tag{2}$$

Das Modell (1) lässt sich durch Transformation mit der Matrix $\boldsymbol{G} = diag\left(\sqrt{p_1},\ldots,\sqrt{p_n}\right)$ in ein Modell mit Einheits-Gewichtsmatrix überführen. Dabei entstehen aufgrund der diagonalen Struktur von $\boldsymbol{G}$ keine zusätzlichen von Null verschiedenen Einträge in der transformierten Koeffizientenmatrix $\boldsymbol{G'X}$. Nachfolgend wird von einem transformierten Modell mit $D(\boldsymbol{y}) = \sigma^2\boldsymbol{I}$ ausgegangen.

3.2 Schätzung der unbekannten Parameter

Die Bestimmung der unbekannten Parameter erfolgt durch ein gedämpftes Gauß-Newton-Verfahren (Hettwer und Benning 2001). In jeder Iteration des Verfahrens wird eine Lösung der Normalgleichungen des linearen Gauß-Markoff-Modells vorgenommen, wobei die mit einem Dämpfungsfaktor $\lambda^{(k)}\,(0<\lambda^{(k)}<1)$ multiplizierten Unbekannten-Zuschläge $\lambda^{(k)}\boldsymbol{\beta}^{(k)}$ der k-ten Iteration als Näherungswerte für die $(k+1)$-te Newton-Iteration benutzt werden.

$$\boldsymbol{X}'^{(k)}\boldsymbol{X}^{(k)}\Delta\boldsymbol{\beta}^{(k)} = \boldsymbol{X}'^{(k)}\Delta\boldsymbol{y}^{(k)} \tag{3}$$

$$\boldsymbol{\beta}^{(k+1)} = \boldsymbol{\beta}^{(k)}\lambda^{(k)}\Delta\boldsymbol{\beta}^{(k)} \tag{4}$$

$$\boldsymbol{X}^{(k+1)} = \left(\left.\frac{\partial h_i(\boldsymbol{\beta})}{\partial\beta_j}\right|_{\beta^{(k+1)}}\right) \tag{5}$$

$$\Delta\boldsymbol{y}^{(k+1)} = \left(y_i - h_i\left(\boldsymbol{\beta}^{k+1}\right)\right) \tag{6}$$

Mit den direkten und den iterativen Gleichungslösern existieren zwei grundlegend verschiedene Ansätze zur Lösung des linearen Normalgleichungssytems (3) (Opfer 1993).

Direkte Verfahren basieren für allgemeine Matrizen auf einer Gauß-Faktorisierung, im Falle positiv-definiter, symmetrischer Matrizen kann das Cholesky-Verfahren zum Einsatz kommen. Liegt eine Faktorisierung der Normalgleichungsmatrix vor, kann mittels Vorwärts-Rückwärtseinsetzen der Lösungsvektor bestimmt werden.

Bei der Verwendung voll besetzter Matrizen wächst der Aufwand zur Lösung des Normalgleichungssystems bei dem direkten Verfahren mit $O(u^3)$, d. h. kubisch mit der Anzahl der unbekannten Parameter. Dabei dominiert die rechnerische Komplexität der Cholesky-Faktorisierung mit $u^3/6$ wesentlichen Operationen (Multiplikationen und Divisionen), wohingegen das Vorwärts-Rückwärtseinsetzen u^2 wesentliche Operationen erfordert. Der Speicherplatzbedarf wächst im Allgemeinen quadratisch mit der Anzahl der Unbekannten.

Bei der Homogenisierung von Massendaten entstehen Normalgleichungen mit mehr als 10^6 Unbekannten, die jedoch nur schwach miteinander verknüpft sind. Eine Lösung kann nach heutigen Maßstäben allein aufgrund des Speicherbedarfs nicht mit den Standardverfahren für voll besetzte Matrizen durchgeführt werden.

Spezielle Verfahren für die Lösung und Inversion linearer Gleichungssysteme, deren Koeffizientenmatrix

nur zu einem geringen Anteil mit von Null verschiedenen Elementen besetzt ist, heißen Sparse-Matrix Techniken. Wie gezeigt wird, hängt die rechnerische Komplexität der verschiedenen Sparse-Matrix Techniken wesentlich von der Anzahl und der Verteilung der von Null verschiedenen Nebendiagonalelemente der Normalgleichungsmatrix ab. Bevor die Sparse-Matrix Techniken genauer erläutert werden, erfolgt eine Analyse der Struktur der Normalgleichungsmatrix der Homogenisierung.

3.3 Struktur der Normalgleichungsmatrix

Die Dimension des Normalgleichungssystems ist identisch mit der Anzahl der unbekannten Parameter $u = k + l$, deren Größenordnung im Wesentlichen durch die Anzahl der Koordinaten der Neupunkte k bestimmt ist.

Im vorliegenden Fall unkorrelierter Beobachtungen setzt sich die Normalgleichungsmatrix additiv aus den Beiträgen der einzelnen Beobachtungsgleichungen $\boldsymbol{x}_i'$ mit $\boldsymbol{X} = |\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n|'$ zusammen zu

$$\boldsymbol{X}'\boldsymbol{X} = \left|\boldsymbol{x}_1\boldsymbol{x}_1' + \ldots + \boldsymbol{x}_n\boldsymbol{x}_n'\right| . \qquad (7)$$

Aus (7) ist ersichtlich, dass von Null verschiedene Nebendiagonalelemente nur dort entstehen, wo mehrere unbekannte Parameter an einer Beobachtungsgleichung beteiligt sind.

- Die Beobachtungsgleichungen der Startsystem-Koordinaten der Neupunkte führen zu Verknüpfungen zwischen den unbekannten Koordinaten des Neupunktes und den fünf Transformationsparametern des Systems (14 Einträge je Neupunkt, 10 je Anschlusspunkt in $\boldsymbol{X}$). Die Parameter der Transformation eines Systems sind daher mit sämtlichen Koordinaten-Unbekannten der Punkte des Systems verknüpft.
- Die Beobachtungsgleichungen für die verschiedenen Typen geometrischer Bedingungen werden auf elementare Bedingungen für zwei bis vier Punkte zurückgeführt und führen daher zu Verknüpfungen zwischen den Unbekannten von maximal zwei bis vier Punkten (vier bis acht Einträge in $\boldsymbol{X}$).
- Die Beobachtungen zum Erhalt der Nachbarschaft werden beim hybriden Ansatz als Differenz-Beobachtungen zwischen den x und y Koordinaten benachbarter Punkte aufgebaut. Die Definition der Nachbarschaft erfolgt mit dem Natural-Neighbour Kriterium. Die Nachbarschaftsbeobachtungen führen zu Verknüpfungen der Koordinaten-Unbekannten solcher Punkte, die innerhalb der Delaunay-Triangulation der Punkte eines Startsystems durch eine Kante verbunden sind (vier Einträge in $\boldsymbol{X}$ je Kante zwischen zwei Neupunkten).
  Die Delaunay-Triangulation ist ein planarer Graph mit n Knoten und besitzt folglich maximal $3n-6$ Kanten. Durch die Nachbarschaftsbeobachtungen werden die Unbekannten eines Neupunktes mit durchschnittlich maximal $2(3n-6)/n \approx 6$ Unbekannten benachbarter Punkte verknüpft.

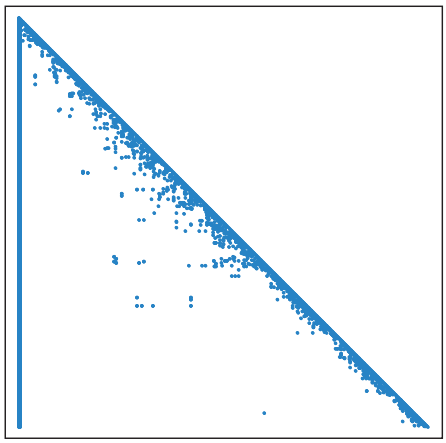

Abb. 4: Nichtnullelemente der Normalgleichungsmatrix des Beispiels aus Abb. 5

In Abb. 4 ist die Struktur der Normalgleichungsmatrix für ein praktisches Homogenisierungsproblem mit ca. 800 Punkten und 160 geometrischen Bedingungen dargestellt. Gut erkennbar sind die dicht besetzten Spalten der Transformationsparameter und die unregelmäßige Verteilung der Nebendiagonalelemente aufgrund der Nachbarschaftsbeobachtungen und der geometrischen Bedingungen.

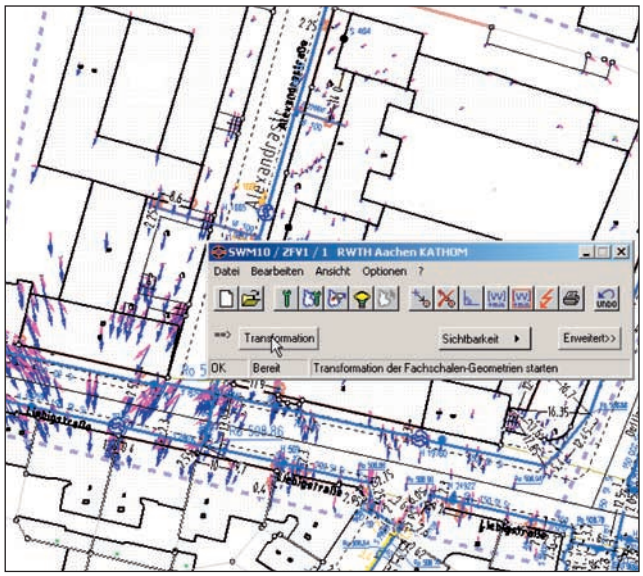

Abb. 5: Homogenisierungsverfahren für die Bestandsdatenaktualisierung eines Netzinformationssystems (hier: Sparte Wasser)

4 Sparse-Matrix Techniken zur Lösung und Inversion des Normalgleichungssystems

Verschiedene Verfahren für die direkte und iterative Lösung und Inversion sparser Normalgleichungssysteme sind seit Jahrzehnten bekannt (Björck 1996, Duff et al. 1986, George und Liu 1981, Schwarz et al. 1968) und kommen seit längerem in vielen geodätischen Anwen-

dungen zum Einsatz (z. B. Ackermann et al. 1970, Benning 1986, Gründig 1976, Stark 1984). Es existieren zahlreiche, zum Teil frei verfügbare Software-Bibliotheken, auf die bei der Implementierung zurückgegriffen werden kann (z. B. Arantes 2004).

Trotz der langjährigen Forschung auf diesem Gebiet ergeben sich fortlaufend neue Erkenntnisse, die zu weiteren Verbesserungen der Verfahren führen. Ein Beispiel hierfür ist das System METIS, das auf einem Multilevel Algorithmus zur Partitionierung unregelmäßiger Graphen basiert (Karypis und Kumar 1998a, Karypis und Kumar 1998b). METIS lässt sich vorzüglich zur Fill-In reduzierenden Sortierung der Normalgleichungsmatrix verwenden. METIS liefert in vielen Fällen, so auch in der Homogenisierung, bessere Ergebnisse als der Generalised Multiple Minimum Degree Algorithmus, der Mitte der 80er-Jahre von Liu entwickelt wurde und aufgrund seiner sehr schnellen Laufzeit weit verbreitet ist (Liu 1985).

Da bereits umfangreiche Literatur zu den klassischen Sparse-Matrix Techniken existiert, ist deren Darstellung hier auf das Wesentliche beschränkt; eine umfangreichere Einführung in die Thematik bieten Schenk und van der Vorst (2004).

4.1 Speicherstrukturen für Sparse-Matrizen

Die speziellen Speicherstrukturen für Sparse-Matrizen sind dahingehend optimiert, dass möglichst wenige Nullelemente gespeichert werden müssen. Die interne Indizierung sparser Strukturen weicht in der Regel ab von der externen Indizierung (Zeile/Spalte) der Matrix; zusätzliche Felder für die Lokalisierung der Elemente müssen vorgehalten werden.

Die Koordinatenspeichertechnik und die Profilspeichertechnik sind die am meisten verbreiteten Klassen von Speicherstrukturen für Sparse-Matrizen. Bei der Koordinatenspeichertechnik werden ausschließlich von Null verschiedene Elemente und deren Koordinaten (Indices) gespeichert, wohingegen bei der Profilspeichertechnik jeweils die Vektoren von der Diagonale bis zum letzten von Null verschiedenen Element, die auch als Profil oder Envelope der Matrix bezeichnet werden, vorgehalten werden. Die Profilspeichertechnik kommt im Vergleich zur Koordinatenspeichertechnik mit einfacheren Lokalisierungsstrukturen aus. Dafür kann das Profil einer Matrix eine erhebliche Anzahl von Nullelementen enthalten, sodass sich ein erhöhter Speicherbedarf der Koordinatenspeichertechnik ergeben kann.

4.2 Direkte Lösungsverfahren

Die Koeffizientenmatrix der Homogenisierung besitzt vollen Spaltenrang. Die Normalgleichungsmatrix $\boldsymbol{X'X}$ der Homogenisierung ist somit positiv-definit. Im Folgenden werden direkte Lösungsverfahren für symmetrische, positiv-definite sparse Systeme linearer Gleichungen betrachtet, die prinzipiell in den drei Phasen *Analyse*, *Faktorisierung* und *Lösung* ablaufen (Duff et al. 1986).

1. Analyse der Normalgleichungsmatrix mit dem Ergebnis einer Fill-In reduzierenden Sortierung der Unbekannten und anschließende symbolische Faktorisierung.
2. Numerische Faktorisierung der Normalgleichungsmatrix $\boldsymbol{X'X = GG'}$ in eine obere und untere Dreiecksmatrix mit Hilfe des Cholesky-Algorithmus für positiv-definite Matrizen.
3. Vorwärts-Rückwärtseinsetzen zur Lösung des Unbekanntenvektors β.

4.2.1 Sortierung der Unbekannten zur Reduzierung der Fill-In Elemente

Die Cholesky-Faktorisierung einer Sparse-Matrix $\boldsymbol{X'X}$ führt in der Regel zur Entstehung zusätzlicher Nichtnullelemente in der Faktormatrix $\boldsymbol{G}$, deren Gesamtheit als Fill-In bezeichnet wird. Die Größe des Fill-Ins ist abhängig von der Reihenfolge der Unbekannten, sodass durch eine Umsortierung der Unbekannten das Fill-In reduziert werden kann (vgl. Abb. 6). Eine Reduzierung des Fill-Ins führt schließlich zu einer Verringerung des Speicherbedarfs und der Anzahl der notwendigen wesentlichen Operationen.

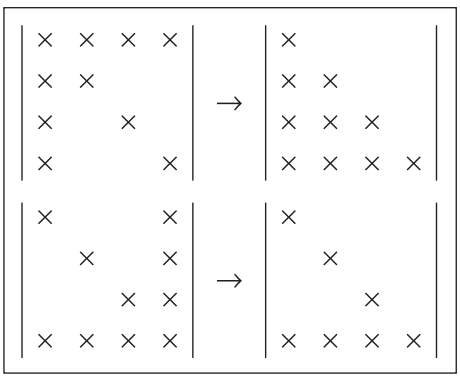

Abb. 6: Effekt der Unbekannten-Sortierung (links) auf die Faktormatrix (rechts) (Duff et al. 1986)

Das Problem der Fill-In reduzierenden Sortierung der Diagonalelemente einer Sparse-Matrix ist NP-vollständig. Nach heutigem Kenntnisstand existiert kein in polynomieller Zeit ablaufender Algorithmus zur Ermittlung einer optimalen Sortierung (Yannakakis 1981). Es gibt eine Reihe heuristischer Verfahren zur Fill-In Reduktion, die jedenfalls für die Praxis brauchbare Ergebnisse in kürzerer Zeit liefern.

Im Rahmen dieser Untersuchung wurden drei verschiedene Gruppen von Sortieralgorithmen miteinander verglichen.

Banker-Sortierung

Die *Banker-Sortierung* wird im Zusammenhang mit der Profilspeichertechnik eingesetzt. Sie führt zu einer Verkürzung der Profilvektoren und somit zu einer Reduktion des benötigten Speicherplatzes (Snay 1976), da Fill-In

Elemente ausschließlich innerhalb des Profils entstehen (Björck 1996, S. 218). Das Verfahren ist in geodätischen Anwendungen weit verbreitet und wird beispielsweise in den Programmsystemen KAFKA und FLASH des Geodätischen Instituts der RWTH Aachen verwendet.

Minimum Degree Algorithmus

Der *Minimum Degree* Algorithmus verfolgt einen intuitiven Ansatz zur Minimierung des Fill-Ins. Aus dem Beispiel in Abb. 6 wird ersichtlich, dass Unbekannte mit einer hohen Zahl von Verknüpfungen zu anderen Unbekannten (man spricht auch vom Grad *(degree)* einer Unbekannten) tendenziell weniger Fill-In Elemente erzeugen, wenn sie an das Ende des Unbekanntenvektors sortiert werden. Die Elimination einer Unbekannten β_i führt zu einer paarweisen Verknüpfung aller mit β_i verknüpften Unbekannten durch Fill-In Elemente, soweit diese nicht schon vor der Elimination von β_i verknüpft waren. Der Minimum Degree Algorithmus wählt nun in jedem Eliminationsschritt die Unbekannte mit minimalem Grad zur folgenden Elimination aus.

Der Hauptaufwand des Minimum Degree Algorithmus liegt in der Aktualisierung des Grades der Nachbarn eliminierter Unbekannter. Weiterentwicklungen setzen an dieser Schwachstelle an. Zu nennen sind Lius *Generalised Multiple Minimum Degree* Algorithmus *(MMD)*, der jeweils eine Menge von Unbekannten in einem Schritt eliminiert, sowie der *Approximate Minimum Degree* Algorithmus *(AMD)*, bei dem eine rechnerisch günstige Näherung des Grades der Unbekannten ermittelt wird (Liu 1985, Davis et al. 1996).

Besondere Schwierigkeiten bereiten Unbekannte mit sehr hohem Verknüpfungsgrad, wie z. B. die Transformationsparameter $\boldsymbol{\beta}_t$, die mit sämtlichen anderen Unbekannten eines Systems verknüpft sind. Sie verursachen bei den Minimum Degree Ansätzen hohen Rechenaufwand, da der Grad dieser Unbekannten bei jedem Eliminationsschritt neu bestimmt werden muss. Unbekannte mit überdurchschnittlich hohem Verknüpfungsgrad sollte man daher vorab von der Sortierung ausschließen und an das Ende der Diagonalen der Normalgleichungsmatrix positionieren. Eine entsprechende Modifikation des AMD Algorithmus ist der *AMDpre* Algorithmus nach (Carmen 1997), der a priori eine Unterteilung des Normalgleichungssystems (3) in ein sparses Teilsystem für $\boldsymbol{\beta}_z$ und ein dicht besetztes System für $\boldsymbol{\beta}_t$ vornimmt.

$$\begin{vmatrix} \boldsymbol{X}_1'\boldsymbol{X}_1 & \boldsymbol{X}_1'\boldsymbol{X}_2 \\ \boldsymbol{X}_2'\boldsymbol{X}_1 & \boldsymbol{X}_2'\boldsymbol{X}_2 \end{vmatrix} \begin{vmatrix} \boldsymbol{\beta}_z \\ \boldsymbol{\beta}_t \end{vmatrix} = \begin{vmatrix} \boldsymbol{X}_1'\boldsymbol{y} \\ \boldsymbol{X}_2'\boldsymbol{y} \end{vmatrix} \tag{8}$$

Im Rahmen dieser Untersuchung wurde Carmens Präprozessor zur Berücksichtigung dicht besetzter Spalten kombiniert mit Lius MMD Algorithmus, das Resultat nennen wir *MMDpre*.

Unabhängig von der Sortierung kann die Aufteilung des Systems (8) auch für den gesamten Lösungsprozess vollzogen werden. Mit $\boldsymbol{N}_{ij} = \boldsymbol{X}_i'\boldsymbol{X}$ erhält man durch Elimination von $\boldsymbol{\beta}_z$ aus (8) die Gleichungen

$$\left(\boldsymbol{N}_{22} - \boldsymbol{N}_{21}\boldsymbol{N}_{11}^{-1}\boldsymbol{N}_{12}\right)\boldsymbol{\beta}_t = \left(\boldsymbol{X}_2' - \boldsymbol{N}_{21}\boldsymbol{N}_{11}^{-1}\boldsymbol{X}_1'\right)\boldsymbol{y} \tag{9}$$

$$\boldsymbol{N}_{11}\boldsymbol{\beta}_z = \boldsymbol{X}_1'\boldsymbol{y} - \boldsymbol{N}_{12}\boldsymbol{\beta}_t\,. \tag{10}$$

Anstelle der Inversen der $(k \times k)$ Sparse-Matrix $\boldsymbol{N}_{11}$ in (9) ermittelt man das Produkt $\boldsymbol{B} = \boldsymbol{N}_{11}^{-1}\boldsymbol{N}_{12}$ als Lösung des Gleichungssystems

$$\boldsymbol{N}_{11}\boldsymbol{B} = \boldsymbol{N}_{12} \tag{11}$$

mit l rechten Seiten. Die Faktorisierung von $\boldsymbol{N}_{11}$ wird auch zur Lösung von (10) benötigt, sodass der zusätzliche Aufwand zur Lösung von (11) in dem l-fachen Vorwärts-Rückwärtseinsetzen mit jeweils k^2 wesentlichen Operationen besteht.

Nested Dissection Algorithmus

Die Nested Dissection Algorithmen basieren auf der Separierung der Unbekannten durch Auswahl einer kleinen Teilmenge von Unbekannten $\boldsymbol{\beta}_s$ (Separator), die den Unbekanntenvektor $\boldsymbol{\beta}$ in zwei unverknüpfte Komponenten $\boldsymbol{\beta}_1$ und $\boldsymbol{\beta}_2$ partitioniert. Die Partitionierung wird innerhalb der Komponenten rekursiv fortgesetzt, bis eine vorgegebene Mindestgröße erreicht wird.

Die Anwendung des Nested Dissection Verfahrens auf die Ausgleichung geodätischer Punktnetze entspricht der Helmert'schen Blockmethode (Helmert 1880, S. 556; Björck 1996, S. 240; Wolf 1975, S. 99). Die Unbekannten $\boldsymbol{\beta}_s$ des Separators werden in dem Zusammenhang Verbindungsunbekannte genannt, die verbleibenden Unbekannten heißen Innenunbekannte.

Das Nested Dissection Verfahren führt auf eine Normalgleichungsmatrix mit Block-Diagonal-Struktur, wobei die Größe des Randes (Puffersystem) abhängig ist von der Größe der Separatoren. Die Block-Diagonal-Struktur ist invariant bzgl. der Cholesky-Faktorisierung (Björck 1996, S. 225); somit hängt die Fill-In Reduktion des Nested Dissection Verfahrens direkt von der Größe der Separatoren ab.

Das Problem der Separierung einer Menge von Unbekannten in zwei unverknüpfte Komponenten kann auch als Partitionierung des Adjazenzgraphen der Matrix interpretiert werden (Friedrich 1930). Die Knotenmenge V mit $|V| = n$ des Adjazenzgraphen $G = (V, E)$ ist gegeben durch die Elemente des Unbekanntenvektors $\boldsymbol{\beta}$; die Kanten E von G sind definiert durch die von Null verschiedenen Nebendiagonalelemente (Verknüpfungselemente) der Normalgleichungsmatrix $\boldsymbol{X}'\boldsymbol{X}$.

Das Problem der Graphenpartitionierung besteht darin, zwei Teilmengen V_1 und V_2 gleicher Größe $|V_1| = |V_2| = n/2$ mit leerer Schnittmenge $\mathrm{V1} \cap \mathrm{V2} = \emptyset$ und vollständiger Vereinigung $V_1 \cup V_2 = V$ zu finden, sodass die Menge der Kanten aus E, deren Knoten zugleich V_1 und V_2 angehören, minimiert wird. Die Graphenparti-

tionierung ist offenbar äquivalent zur Suche eines guten Separators für das Nested Dissection Verfahren.

Eine effiziente Heuristik zur Lösung des Graph-Partitionierungs-Problems im Kontext der Matrix-Sortierung ist durch den Kernighan-Lin Algorithmus gegeben (Kernighan und Lin 1970). Ausgehend von initialen Partitionierungen des Adjazenzgraphen sucht man Teilmengen von Knoten, deren Vertauschung zwischen den Partitionen zu einer Verkleinerung des Separators führt. Das Verfahren wird iterativ fortgesetzt bis ein lokales Minimum erreicht wird und keine entsprechenden Teilmengen von Knoten mehr gefunden werden.

Eine weitere Beschleunigung der Graphenpartitionierung gewährleistet die Multilevel Technik, die auf den Multilevel Nested Dissection (MLND) Algorithmus führt. Der MLND Algorithmus verläuft in drei Phasen, der Zusammenfassungsphase (coarsening phase), der Partitionierungsphase und der Verfeinerungsphase (uncoarsening phase) (vgl. Abb. 7). Wie bei Multilevel Algorithmen üblich, wird zunächst eine stufenweise Vereinfachung der

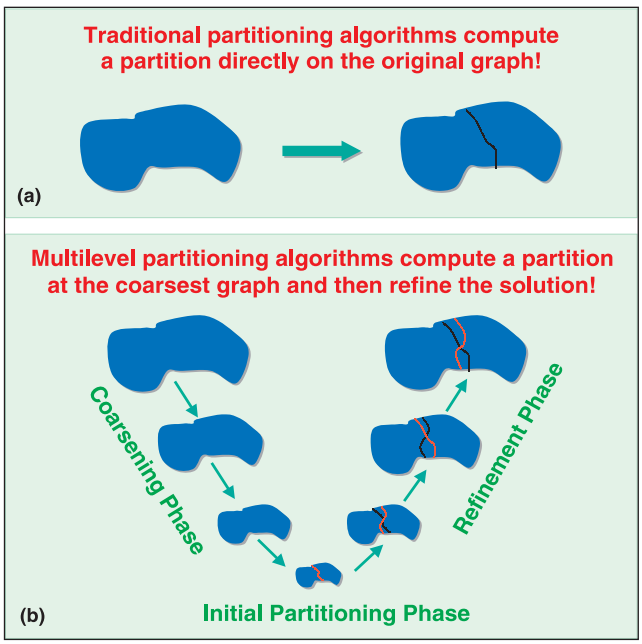

Abb. 7: Vorgehensweise bei traditioneller (a) und Multilevel Partitionierung (b) [Karypis und Kumar 1998b]

Problemstellung, hier durch Reduktion der Dimension des Graphen G_0 durch eine Zusammenfassung von Knoten zu Hyperknoten in mehreren Stufen $G_1, \ldots, G_m$ durchgeführt. Es folgt eine Partitionierung P_m des maximal vereinfachten Graphen G_m und schließlich wird die Partitionierung P_m von G_m zurückprojiziert bis zum Graphen G_0. Dabei werden die zwischenliegenden Partitionen $P_{m-1}, \ldots, P_0$ sukzessive durchschritten und es erfolgt jeweils eine lokale Verbesserung der projizierten Partitionen mit dem Kernighan-Lin-Algorithmus (Karypis und Kumar 1998a).

Die Unbekannten mit hohem Verknüpfungsgrad sollten beim MLND Algorithmus getrennt behandelt werden, um eine optimale Performanz zu erzielen.

Implementierung im Programm KATHOM

Sämtliche der beschriebenen Verfahren sind in das Programmsystem KATHOM integriert worden. Die Banker-Sortierung wurde aus dem Vorgängerprogramm FLASH übernommen (Benning und Scholz 1990). Der Sortieralgorithmus *DSS_REORDER* aus der Compaq Extended Math Library, dessen theoretischer Hintergrund nicht publiziert ist, wurde von Hettwer integriert (Compaq Computer Corporation 2001, Hettwer 2003).

Neu sind der Generalised Multiple Minimum Degree Algorithmus, der Approximate Minimum Degree Algorithmus und das Multilevel Nested Dissection Verfahren aus METIS. Dabei wurden jeweils Algorithmen zur Berücksichtigung der Unbekannten mit hohem Verknüpfungsgrad vorgeschaltet.

4.2.2 Faktorisierung und Lösung sparser Normalgleichungssysteme

Die Fill-In reduzierende Sortierung der Unbekannten, die allein aufgrund der Verknüpfungsinformation der Normalgleichungsmatrix durchgeführt wird, liefert einen Permutationsvektor der eine neue Sortierung der Unbekannten vorgibt. Auf der Grundlage dieser Information erfolgt im nächsten Schritt die Faktorisierung der Matrix, wobei zunächst eine symbolische Faktorisierung der permutierten Matrix durchgeführt wird. Auf deren Grundlage werden die Positionen der Fill-In Elemente berechnet und die Felder der Koordinatenspeichertechnik aufgebaut. Im Anschluss daran erfolgt die numerische Faktorisierung, für die sich der *left-looking* Algorithmus und die *multifrontal* Methode anbieten (Schenk und van der Vorst 2004). Weiterhin existieren einige Ansätze zur Optimierung der numerischen Faktorisierungsverfahren, die der hierarchischen Speicherstruktur moderner Rechner angepasst sind.

Die Anwendung des Nested Dissection Verfahrens führt offenbar auf ein Normalgleichungssystem mit Block-Diagonal-Struktur, für dessen parallelisierte Lösung sich die gruppenweise Ausgleichung nach dem Zusammenschließungstyp unter Anwendung von Helmerts Additionstheorem für reduzierte Normalgleichungen anbietet (Wolf 1975, S. 100, 114).

Implementierung im Programm KATHOM

Im Programm KATHOM wurden die Algorithmen aus George und Liu (1981) zur symbolischen und numerischen Faktorisierung und für das Vorwärts-Rückwärtseinsetzen verwendet, die den *left-looking* Ansatz verwenden und keine numerischen Pivot-Strategien enthalten. Die Normalgleichungsmatrix der Homogenisierung ist stets positiv-definit, sodass keine singulären führenden Hauptuntermatrizen auftreten. Eine Pivotisierung oder dazu alternative Vorgehensweise ist nicht erforderlich (Koch 1997, S. 28, 66, 202). Eine Parallelisierung der Lösung wurde testweise vorgenommen.

4.2.3 Partielle Berechnung der Kovarianzmatrix der unbekannten Parameter

Zur Berechnung des Parametervektors β aus den Normalgleichungen (3) ist eine Inversion der Normalgleichungsmatrix nicht erforderlich. Allerdings werden die Elemente der Inversen $(X'X)^{-1}$ zur Berechnung der Kovarianzmatrix $\hat{D}(\hat{\beta}) = \hat{\sigma}^2(X'X)$ benötigt, deren Ermittlung beispielsweise zur Durchführung von Hypothesentests erforderlich ist.

Die Inverse einer Sparse-Matrix ist im Allgemeinen voll besetzt, sodass eine vollständige Berechnung aller Elemente der Inversen bei der vorliegenden Problemstellung nicht in Frage kommt. Allerdings werden zur Angabe der a posteriori Varianz einzelner Parameter sowie zur Durchführung von Hypothesentests nur ausgewählte Elemente der Inversen-Matrix benötigt.

Ein effizienter Algorithmus zur partiellen Inversion von Sparse-Matrizen wurde Anfang der 70er-Jahre von Takahashi angegeben (Takahashi et al. 1973, Erisman und Tinney 1975). Das Verfahren basiert auf der Gauß-Zerlegung einer hier symmetrischen $n \times n$ Matrix $A = LDL'$, für deren Inverse $Z = A^{-1}$ die folgende Beziehung gilt:

$$Z = L'^{-1} D^{-1} L^{-1} \tag{12}$$

$$= L'^{-1} D^{-1} L^{-1} + \overbrace{(D^{-1}L^{-1} - D^{-1}L^{-1})}^{0} \tag{13}$$

$$= D^{-1}L^{-1} + (I - L')Z \quad . \tag{14}$$

Aus der Beziehung (14) lässt sich eine Rekursionsvorschrift für die Elemente der Inversen $Z = (z_{ij})$ ableiten. Ausgehend vom Element $z_{nn} = d_{nn}^{-1}$ erhält man die weiteren Elemente von Z aus den Zusammenhängen

$$z_{ij} = -\sum_{k=i+1}^{n} l_{ki} z_{kj} \tag{15}$$

$$z_{ii} = d_{nn}^{-1} - \sum_{k=i+1}^{n} l_{ki} z_{ki}$$
$$\text{mit } i = n-1,\ldots,1 \text{ und } j > i. \tag{16}$$

Eine für die Inversion von Sparse-Matrizen besonders günstige Eigenschaft des Algorithmus besteht darin, dass die Berechnung der Elemente z_{ij} mit $l_{ij} \neq 0$ ohne vorherige Berechnung der Elemente z_{ij} mit $l_{ij} = 0$ erfolgen kann. Werden ausschließlich solche Elemente der Inversen benötigt, die bereits in A von Null verschieden sind oder als Fill-In Elemente in der Faktorisierung hinzugekommen sind, kann die Berechnung der Inversen A^{-1} ohne zusätzlichen Speicherbedarf geschlossen innerhalb der Sparse-Matrix Struktur der Faktormatrix $G' = \sqrt{D}L'$ erfolgen.

Implementierung im Programm KATHOM

Der Takahashi-Algorithmus zur Berechnung der partiellen Inversen ist in das Programm KATHOM integriert worden, wobei nur die in der Struktur der Cholesky-Faktorisierung enthaltenen Elemente berechnet werden. Weitere Elemente der Inversen werden nicht benötigt, denn die im weiteren Verlauf zur Durchführung der Ausreißertests benötigten Elemente sind bereits in den Normalgleichungen von Null verschieden.

5 Anwendungsbeispiele

In den Abb. 8 und 9 sind die Ergebnisse der verschiedenen Sortierverfahren für das Testbeispiel aus Abb. 5 dargestellt, die Ergebnisse des AMD sind aufgrund der Ähnlichkeit zum MMD nicht dargestellt. Bei dem Multi Level Nested Dissection Ansatz ist die Partitionierung des Problems in jeweils $(200 \times 200)/2$ Elemente große Blöcke erkennbar, die abgesehen von den Separatoren und deren Fill-In auch in der Faktormatrix erhalten bleibt. Diese Struktur macht sich bei Verwendung des left-looking Algorithmus zur numerischen Faktorisierung der Matrix positiv bemerkbar, da innerhalb eines lokal zusammenhängenden Abschnitts des Elementvektors gearbeitet werden kann, der vollständig in den Cache-Speicher der CPU passt. So ist zu erklären, dass die Faktorisierung und Lösung der mit MLND sortierten Matrizen schneller abläuft als die Faktorisierung der MMD- und AMD-Matrizen, obwohl diese zum Teil ein kleineres Fill-In besitzen.

Die Laufzeiten und die Größe des Fill-Ins der verschiedenen Sortieralgorithmen sind für einige Datensätze aus

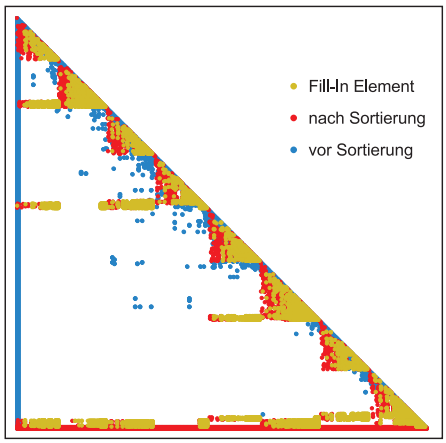

Abb. 8: Ergebnisse des MLND Algorithmus (METIS)

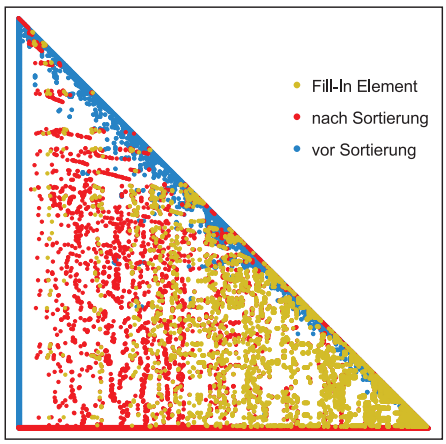

Abb. 9: Ergebnisse des MMD Algorithmus

Tab. 1: Testdatensätze mit bis zu 1,16 Millionen Unbekannten

	Neup.	Sollp.	Geom. Bed.	Nebendiag. (%)
1	18002	14284	20780	430663 (0,069)
2	143392	14863	27715	2517627 (0,006)
3	493088	115413	80057	8562485 (0,002)
4	578727	115286	199335	10779796 (0,002)

realen Homogenisierungsproblemen (vgl. Tab. 1) in den Abb. 10 und 11 dargestellt. Die Rechenzeiten wurden auf einem AMD Athlon 2800+ mit 1 Gb RAM ermittelt. Der MLND Algorithmus ist geringfügig langsamer als der MMDpre Algorithmus. Die vergleichsweise hohen Laufzeiten der DSS-CXML Unbekannten-Sortierung sind auf eine Unzulänglichkeit bei der Behandlung von Unbekannten mit hohem Verknüpfungsgrad zurückzuführen, die man beispielsweise mit der Vorgehensweise nach Gleichung (9)–(11) umgehen kann.

Misst man die Qualität eines Sortierverfahrens anhand der Laufzeit der anschließenden numerischen Faktorisierung, so ist aus Abb. 12 erkennbar, dass MLND (METIS) den anderen Verfahren aufgrund des schnelleren Speicherzugriffs auch dann überlegen ist, wenn bei größerem

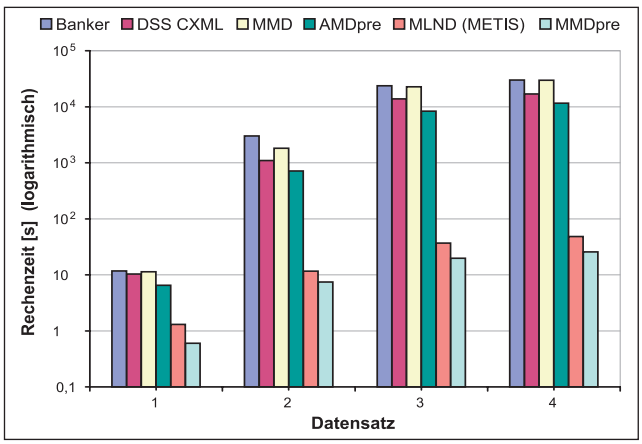

Abb. 10: Rechenzeiten der Unbekannten-Sortierung

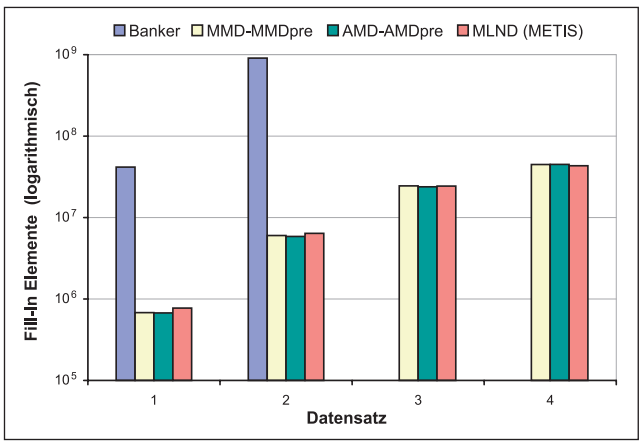

Abb. 11: Anzahl der Fill-In Elemente (bei DSS-CXML nicht zugänglich)

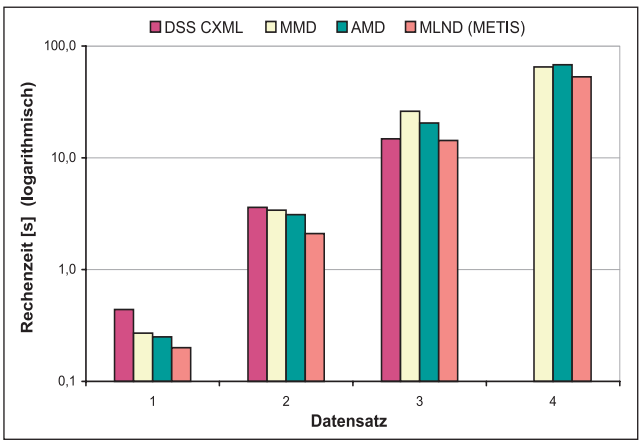

Abb. 12: Rechenzeiten der numerischen Faktorisierung und Lösung

Fill-In eine größere Zahl von wesentlichen Operationen benötigt wird. Für einige Datensätze fehlen Angaben zur Laufzeit der numerischen Faktorisierung nach Banker-Sortierung und DSS-CXML, da die 2Gb-Schranke für Benutzerprozesse in 32-bit Windows-Systemen überschritten wurde.

In Kombination mit MLND ist eine Steigerung der Rechenleistung durch den Einsatz parallelisierter Faktorisierungs- und Lösungsverfahren auf Multiprozessor-Rechnern erzielbar (vgl. 4.2.2) (Gupta et al. 1997, Karypis und Kumar 1998a, Schenk 2000). Vielversprechende Testläufe mit den Datensätzen aus Tab. 1 wurden mit dem parallelen direkten Gleichungslöser PARDISO (Version 1.2.2) auf einem IBM Power 4 mit vier 1.3 GHz CPUs der Uni Basel durchgeführt. Die Parallelisierung auf vier Prozessoren bewirkt für Datensatz 4 eine Beschleunigung der numerischen Faktorisierung um den Faktor 3,5 von 15,5 s auf 4,4 s.

Die partielle Inversion der Normalgleichungsmatrix erfordert in den Datenbeispielen näherungsweise doppelt so viele wesentliche Operationen wie die numerische Faktorisierung und Lösung der Normalgleichungen. Die Inverse wird nur einmal am Ende der Gauß-Newton-Iterationen berechnet, sodass sie nur unwesentlich zur Gesamtrechenzeit der Homogenisierung beiträgt. Zur Qualitätsbeurteilung und Durchführung statistischer Tests sollte die partielle Inverse (Kovarianzmatrix) daher immer berechnet werden.

In Abb. 13 sind die Gesamtlaufzeiten der Programme KATHOM 1.10 (Banker-Algorithmus), KATHOM 1.18 (DSS-CXML) und KATHOM 2.0 (MLND) aufgeführt. Der Datensatz 4 konnte aufgrund des erhöhten Speicherbedarfs der anderen Verfahren nur mit der Version 2 des Programms berechnet werden. Bei der Bewertung der Laufzeiten ist zu berücksichtigen, dass die in Version 1 verwendeten Algorithmen zur Delaunay-Triangulation und Natural Neighbour Interpolation eine Punkt-in-Umkreis-Suche mit linearem Aufwand verwenden, wodurch ein insgesamt quadratisches Laufzeitverhalten vorliegt (Hettwer 2003, S. 95). In der Version 2 wird ein räumlicher Index benutzt, der zumindest für zufällig gleich-

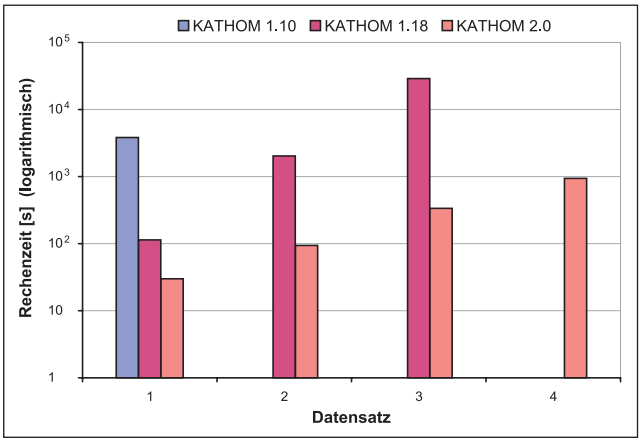

Abb. 13: Gegenüberstellung der Gesamtlaufzeiten der Homogenisierung

verteilte Punktmengen eine Lokalisierung in konstanter Zeit zulässt (Tsai 1993).

Abb. 14 stellt die Verbesserungsbeträge der Homogenisierung eines Teilgebietes der Münchener Innenstadt dar, die lokale Inhomogenität der Daten ist erkennbar.

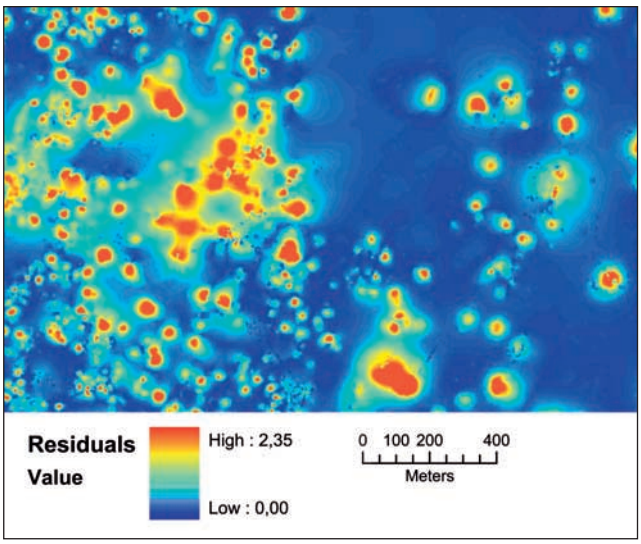

Abb. 14: Residuenbeträge der Homogenisierung eines Teilgebietes der Münchener Innenstadt

6 Bewertung und Ausblick

Die praxisrelevante Aufgabe der Homogenisierung von Massendaten wurde untersucht und es wurden erhebliche Verbesserungen gegenüber den bisher im Programmsystem KATHOM angewandten Techniken aufgezeigt. Auch im klassischen Gebiet der numerischen linearen Algebra, das seit über einem halben Jahrhundert Gegenstand der Forschung ist, ist stets mit neuen, für die Rechenverfahren der Geodäsie relevanten Erkenntnissen zu rechnen.

Der Nutzen dieser wissenschaftlichen Evolution macht sich für den Praktiker unmittelbar bezahlt. Beispielsweise kann die Homogenisierung einschließlich Inversion von 25 % der Bestandsdaten des Netzinformationssystems der Münchener Innenstadt (1.2 Millionen Unbekannte) auf einem handelsüblichen 32-bit Rechner mit KATHOM 2.0 in weniger als 20 Minuten erfolgen.

Die in Abschnitt 2 dargestellten neuen Anwendungsbereiche im Kontext der Geodaten-Infrastrukturen zeigen, dass weitere Herausforderungen im Bereich der Homogenisierung von Massendaten anstehen.

Der Einsatz parallelisierter Algorithmen auf 64-bit Multiprozessor-Rechnern wird zu einer weiteren Steigerung der Kapazität und Leistungsfähigkeit der Homogenisierung führen.

Dank

Unser Dank gilt der Stadtwerke München GmbH für die freundliche Freigabe der Daten zur Veröffentlichung in diesem Beitrag. Wir danken Herrn Dr. Olaf Schenk, Department of Computer Science, Uni Basel, für seine Diskussionsbereitschaft, den Hinweis zur Möglichkeit der Sparse-Elimination nach (11) und das Testen einiger Datenbeispiele mit PARDISO auf den Parallelrechnern der Uni Basel.

Literatur

Ackermann, F., Ebner, H., Klein, H.: Ein Programm-Paket für die Aerotriangulation mit unabhängigen Modellen. Bildmessung und Luftbildwesen, 38: 218-224, 1970.

Arantes, R. D.: Information Base on Sparse Linear Systems and Matrix Computations. http://openlink.br.inter.net/sparse/main.html, 2004.

Benning, W.: Analyse hybrider Lageaufnahmen in Sparse-Technik. Zeitschrift für Vermessungswesen, 111: 506-513, 1986.

Benning, W.: Nachbarschaftstreue Restklaffenverteilung für Koordinatentransformationen. Zeitschrift für Vermessungswesen, 120: 16-25, 1995.

Benning, W.: Zur Homogenisierung und Fortführung von großmaßstäblichen Karten. Vermessung, Photogrammetrie, Kulturtechnik, 2: 75–80, 1996.

Benning, W., Scholz, T.: Modell und Realisierung der Kartenhomogenisierung mit Hilfe strenger Ausgleichungstechniken. Zeitschrift für Vermessungswesen, 115: 44-55, 1990.

Bernard, L., Einspanier, U., Lutz, M., Portele, C.: Interoperability in GI Service Chains – The Way Forward. In: Proccedings of the 6th AGILE, Lyon, 2003.

Bishr, Y.: Overcoming the Semantic and Other Barriers to GIS Interoperability. International Journal of Geographical Information Science, 12(4): 299-314, 1998.

Björck, A.: Numerical Methods for Least Squares Problems. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa. 1996.

Carmen, J. L.: An AMD Preprocessor for Matrices with Some Dense Rows and Columns. Technical Report, University of Florida, Department of Computer and Information Science Engineering, Gainesville, Fl. 1997.

Compaq Computer Corporation: Compaq Extended Math Library Reference Guide. Houston, Tx. 2001.

Conrad, S.: Föderierte Datenbanksysteme – Konzepte der Datenintegration. Springer, Berlin, 1997.

Davis, T. A., Amestoy, P., Duff, I.: An Approximate Minimum Degree Ordering Algorithm. SIAM Journal on Matrix Analysis and Applications, 113: 581–587, 1996.

Duff, I. S., Erisman, A. M., Reid, J.: Direct Methods for Sparse Matrices. Clarendon Press, Oxford, 1986.

Erisman, A. M., Tinney, W. F.: On Computing Certain Elements of the Inverse of a Sparse Matrix. ACM Communications, 18: 177–179, 1975.

Friedrich, K.: Beiträge zur direkten und indirekten Auflösung der Normalgleichungen. Zeitschrift für Vermessungswesen, 59: 525–539, 1930.
George, A., Liu, J.: Computer Solution of Large Sparse Positive Definite Systems. Prentice-Hall, Englewood Cliffs, 1981.
Gielsdorf, F., Gründig, L., Aschoff, B.: Geo-Referencing of Analogue Maps with Special Emphasis on Positional Accuracy Improvement Updates. In: FIG Working Week 2003. International Federation of Surveyors, Paris, France, 2003.
Gielsdorf, F., Gründig, L., Aschoff, B.: Positional Accuracy Improvement – A Necessary Tool for Updating and Integration of GIS Data. In: FIG Working Week 2004. International Federation of Surveyors, Athens, Greece, 2004.
Gröger, G.: Modellierung raumbezogener Objekte und Datenintegrität in GIS. Wichmann, Karlsruhe, 2000.
Gröger, G., Kolbe, T. H.: Interoperabilität in einer 3D-Geodateninfrastruktur. In: Bernard, L., Sliwinski, A., Senkler, K. (Hrsg.): IfGI Prints: Beiträge zu den Münsteraner GI-Tagen 2003, Heft Nummer 18. Institut für Geoinformatik, Uni Münster, 325–343, 2003.
Gründig, L.: Berechnung vorgespannter Seil- und Hängenetze unter Berücksichtigung ihrer topologischen und physikalischen Eigenschaften und der Ausgleichungsrechnung. Reihe C, Heft Nummer 216, DGK, München. 1976.
Gupta, A., Karypis, G., Kumar, V.: Highly Scalable Parallel Algorithms for Sparse Matrix Factorization. IEEE Transactions on Parallel and Distributed Systems, 8 (5): 502–520, 1997.
Hake, G., Grünreich, D., Meng, L.: Kartographie. de Gruyter, Berlin, 8. Ausgabe, 2002.
Hampp, D.: Digitalisierung, Homogenisierung und numerische Fortführung – ein schneller Weg zur aktuellen Digitalen Flurkarte. Mitteilungsblatt DVW Bayern, (3): 239–252. 1. Digitalisierung und Homogenisierung. 1992.
Helmert, F. R.: Die mathematischen und physikalischen Theorien der höheren Geodäsie, Band 1, Teubner, Leipzig, 1880.
Hettwer, J.: Numerische Methoden zur Homogenisierung großer Geodatenbestände. Veröffentlichungen des Geodätischen Instituts der RWTH Aachen, Heft Nummer 60, Geodätisches Institut der RWTH Aachen, 2003.
Hettwer, J., Benning. W.: Nachbarschaftstreue Koordinatenberechnung in der Kartenhomogenisierung. Allgemeine Vermessungs-Nachrichten, 107: 194–197, 2000.
Hettwer, J., Benning. W.: Erweiterung des Konvergenzbereichs bei nichtlinearen Ausgleichungsaufgaben. Allgemeine Vermessungs-Nachrichten, 108: 254–259, 2001.
Illert, A.: Aspekte der Zusammenführung digitaler Datensätze unterschiedlicher Quellen. In: Nachrichten aus dem Karten- und Vermessungswesen, Heft Nummer 113 der Reihe I. Verlag des Instituts für Angewandte Geodäsie, Frankfurt am Main, 1995.
Kampshoff, S., Benning, W.: Integrierte Verarbeitung der Daten des Liegenschaftskatasters einschließlich Homogenisierung. zfv – Zeitschrift für Geodäsie, Geoinformation und Landmanagement, 127: 9–18, 2002.
Karypis, G., Kumar, V.: A Fast and High Quality Multilevel Scheme for Partitioning irregular Graphs. SIAM Journal on Scientific Computing, 20: 359–392, 1998a.
Karypis, G., Kumar, V.: Multilevel Algorithms for Multi-Constraint Graph Partitioning. Technical Report MN 98-019, University of Minnesota, Department of Computer Science, Minneapolis, Mn. 1998b.
Kernighan, B. W., Lin, S.: An Efficient Heuristic Procedure for Partitioning Graphs. The Bell System Technical Journal, 1970.
Koch, K.-R.: Parameterschätzung und Hypothesentests in linearen Modellen. Dümmler, Bonn, 1997.
Kottman, C.: White Paper on Data Fusion. Technischer Bericht, Open GIS Consortium. http://www.opengis.org, 2000.
Laurini, R.: Multi-Source Updating and Fusion of Geographic Databases. Computers, Environment and Urban Systems, 18 (4): 243–256, 1994.
Laurini, R.: Spatial Multi-database Topological Continuity and Indexing: a Step Towards Seamless GIS Data Interoperability. International Journal of Geographical Information Science, 12 (4): 373–402, 1998.
Liesen, L.: ALK- und Netzhomogenisierung mit Homage. Vermessungsingenieur, 4: 199-204, 1998.
Liu, J.: Modification of the Minimum Degree Algorithm by Multiple Elimination. ACM Transactions on Mathematical Software, 11: 141–153, 1985.
Müller, J.: Homogenisierung dreidimensionaler Szenarien nach der Methode der kleinsten Quadrate. Veröffentlichungen des Geodätischen Instituts der RWTH Aachen, Heft Nummer 56, Geodätisches Institut der RWTH Aachen, 1999.
Morgenstern, D., Prell, K.-M., Riemer, H.-G.: Digitalisierung, Aufbereitung und Verbesserung inhomogener Katasterkarten. Allgemeine Vermessungs-Nachrichten, 95 (8–9): 314–324, 1988.
OGC: OpenGIS Reference Model. http://www.opengis.org. 2003.
Opfer, G.: Numerische Mathematik für Anfänger. Vieweg, Braunschweig, 1993.
Riedemann, K., Timm, C.: Services for Data Integration (Special Section on Spatial Data Usability). Data Science Journal, 2: 90–99, 2003.
Rose, A.: Geraden- und Rechtwinkelausgleichung bei der Digitalisierung von Katasterkarten. Zeitschrift für Vermessungswesen, 113: 581–587, 1988.
Schenk, O.: Scalable Parallel Sparse LU Factorisation Methods on Shared Memory Multiprocessors. Dissertation, ETH Zürich, 2000.
Schenk, O., van der Vorst, H. A.: Solution of linear Systems. Erscheint in: Handbook of Numerical Analysis, Elsevier Science Publishers, The Netherlands, 2004.
Scheu, M., Effenberg, W., Williamson, I.: Incremental Update and Upgrade of Spatial Data. Zeitschrift für Vermessungswesen, 4: 115–120, 2000.
Schwarz, H. R., Rutishauser, H., Stiefel, E.: Numerik symmetrischer Matrizen. Teubner, Stuttgart, 1968.
Snay, R. A.: Reducing the profile of sparse symmetric matrices. NOOA Technical Memorandum NOS NGS 4, National Geodetic Survey, Rockville, Md. 1976.
Stark, W.: Untersuchungen zur Lösung und Inversion schwach besetzter großer geodätischer Normalgleichungen. Reihe C, Heft Nummer 301, DGK, München, 1984.
Stockwald, M.: Aktualisierung hybrider geographischer Informationssysteme bei Energieversorgungsunternehmen durch amtliche Geobasisdaten. Reihe C, Heft Nummer 529, Deutsche Geodätische Kommission, München, 2000.
Takahashi, K., Fagan, J., Chen, M. S.: Formation of a Sparse Bus Impedance Matrix and its Application to Short Circuit Study. In: Power Industry Computer Applications Conference, 63–69, 1973.
Tsai, V. J. D.: Delaunay Triangulations in TIN Creation: an Overview and a Linear-Time Algorithm. International Journal of Geographical Information Systems, 7: 501–524, 1993.
Walter, V.: Zuordnung von raumbezogenen Daten – am Beispiel der Datenmodelle ATKIS und GDF. Reihe C, Heft Nummer 480, Deutsche Geodätische Kommission, München, 1997.
Wan, W. Y., Williamson, I.: Problems in Maintaining Associativity in LIS with Particular Reference to the Needs of the Utility Industry. The Austrailian Surveyor, 39 (3): 187–193, 1994a.
Wan, W. Y., Williamson, I.: Solutions to Maintaining Associativity in LIS with Particular Reference to the Needs of the Utility Industry. The Austrailian Surveyor, 39 (4): 290–296, 1994b.
Wiens, H.: Flurkartenerneuerung mittels Digitalisierung und numerischer Bearbeitung unter besonderer Berücksichtigung des Zusammenschlusses von Inselkarten zu einem homogenen Rahmenkartenwerk. Schriftenreihe des Instituts für Kartographie und Topographie der Universität Bonn, Heft Nummer 17, Kirschbaum, Bonn, 1984.
Wolf, H.: Ausgleichungsrechnung, Formeln zur praktischen Anwendung. Dümmler, Bonn, 1975.
Worboys, M. F., Duckham, M.: GIS: A Computing Perspective. CRC Press, Boca Raton, Fl, 2. Ausgabe, 2004.
Yannakakis, M.: Computing the Minimum Fill-In is NP-Complete. SIAM Journal on Algebraic Discrete Methods, 2: 77–79, 1981.

Anschrift der Autoren
Dr.-Ing. Stefan Kampshoff / Univ.-Prof. Dr.-Ing. Wilhelm Benning
Rheinisch-Westfälische Technische Hochschule Aachen
Geodätisches Institut
Templergraben 55, 52056 Aachen
kampshoff@gia.rwth-aachen.de / benning@gia.rwth-aachen.de