

Att forska inom Datavetenskap Computing Science

Kurs: Metoder och verktyg, 2004-09-22

Bo Kågström
Institutionen för
Datavetenskap och HPC2N
Umeå universitet

Datavetenskap

- Brett ämnesområde ...
- Drivande kraft i IT-utvecklingen ...
 - ◆ snabb och delvis oförutsägbar
 - ◆ stor genomslagskraft och omfattande konsekvenser
- För forskningen ...
 - ◆ stor flexibilitet och beredskap att ta itu med nya frågeställningar
- För utbildningen
 - ◆ kursutbud och -innehåll ständigt förnyas

Vad är vetenskaplig forskning?

- Systematiskt arbete vars syfte är att utveckla väsentlig ny kunskap som säkras med hjälp av **matematiska** eller **empiriska metoder** och baseras i den mån det är möjligt på ämnets senaste rön
- Finns flera definitioner:
Stöld (1) → Plagiat (2) → Forskning (>2)
- *Do not reinvent the wheel!*

Vad kännetecknar vetenskap?

Olika förslag till kriterier siktar på olika måltavlor:

- produkten
 - vetenskapliga publikationer, "vetenskaplig kunskap", praktiska resultat...
- institutionen
 - universitet och forskningsorganisation, forskarutbildning och -träning, traditionen, forskarsamhället...
- aktiviteten
 - processen, metod(er)...
- förhållningssättet
 - attityd, mental verktygslåda...

Koppling mellan grund- utbildning och forskning?

Grundutbildningen ska ...

- vilå på vetenskaplig grund
 - ◆ Svag tolkning: meddela vetenskapligt underbyggda resultat
 - ◆ Stark tolkning: meddela också den vetenskapliga underbyggnaden och vari det vetenskapliga består
- vara forskningsankuten
 - ◆ Nya rön tas upp som kursmoment
 - ◆ Projektarbeten (exjobb) utförs i samarbete med en forskargrupp

Datavetenskap-forskning

- Bredrivs inom ett brett spektrum ...
 - ◆ AI-filosofi → ... → programvara → ...
 - formella teoretiska studier ...
- Motiveras såväl av inomvetenskapliga som tillämpningsinriktade frågeställningar
- Flera pågående projekt är mångdisciplinära och fakultetsöverskridande
- Flera internationella samarbetsprojekt (Norden, EU och USA)
- Samarbeten med näringsliv (ABB, IBM, Oryx, Partek, Skanska, m.fl.)

Computing Science

- Intelligenta beräkningar
- Kognitiva beräkningar
- Parallella beräkningar
- Teknisk-vetenskapliga beräkningar
- Programvarukonstruktion
- kravhantering, återanvändning, ...
- Teoretisk datalogi - algebra och logik
- Datavetenskaplig didaktik
- Medicinsk informatik
- Bildbehandling och -analys

Cognitive Computing Lab

- Lars-Erik Janlert, professor
- Anders Broberg, FD, univ.lektor
- Jerry Eriksson, FD, univ.lektor
- Thomas Pedersson, FD

det kognitiva beräkningslaboratoriet studerar de kognitiva och beräkningsmässiga aspekterna av samspelet mellan information, objekt och personer - stödda och utökade med modern informationsteknik

CCL samverkar inom ramen för UCIT

Cognitive Computing Lab

ex forskningsinriktningar:

- Mobilitet** - t.ex. under vilka omständigheter kan oberoende mobila enheter samverka för ett övergripande mål?
- Context Awareness** - automatiskt sammanhangsutnyttjande (exvis position)
- Cognitive Quality of Service** - kvalitet på kommunikation och interaktion mätta ur kognitiv synpunkt: är den god nog för de syften, de uppgifter vi försöker lösa?
- Kognitiva verktyg** - personliga stöd för kunskapsarbete
- Fysisk-virtuella gränssnitt** - överbrygga gapet mellan informationsvärlden och den fysiska vardagsvärlden
- Human-Idea-Object Interaction (HIOI)** - hur ska man integrera olika interaktionsmöjligheter till en fungerande helhet? - utveckla begrepp, teori och metoder för att koordinera och integrera den virtuella informationsvärlden med den fysiska och sociala vardagsvärlden

Exempel aktuella projekt

BASTARD - Bandwidth Adapted Streaming Application Research and Demo - under vilka förutsättningar och i vilka former kan man med behållning ta del av t.ex. en fotbollsmatch på avstånd och trådlöst med en mobil enhet?



Video + animering

eFAIR - Rikare upplevelser, ökat utbyte och ökad omsättning på Mässor - samarbete med NOLIA, ingår i VITAL-samarbetet (UCIT/UmU, II, CBT/LTH)

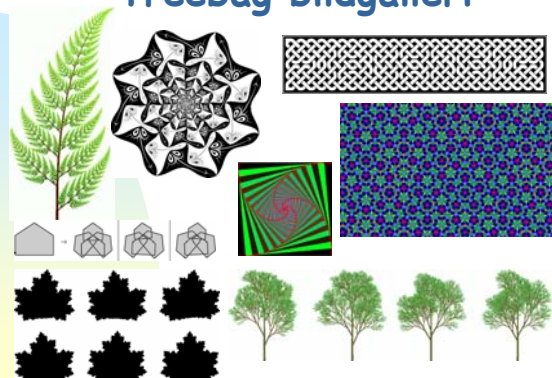
The Character of Things - hur kan man underlätta handhavandet av komplexa system ...



Generering & transformation av träd-, graf- och bildspråk

- Frank Drewes, docent, univ.lektor**
- Regelbaserade system som beskriver mängder av träd, grafer eller bilder samt deras förändring på syntaktisk nivå
- Visuella (program)språk och deras semantik
- "Syntax-directed translation"
- Grammatiker som genererar fraktaler och andra komplexa strukturer
- Programsystemet **Treebag**
- Komplexitet och beräkningsbarhet

Treebag bildgalleri



Intelligenta beräkningar

- Datadriven analys
- Mönsterigenkänning
- Bildanalys
- Adaptiva modeller
- Applikationsnära
- Tvärvetenskapligt
- Beslutsstöd
 - ◆ Medicinsk diagnostik
 - ◆ Finansiell matematik

IFOR

Intelligenta Fordon Off-Road

- Thomas Hellström, FD, univ. lektor

Utveckling av semi-autonoma och autonoma fordon för användning inom

- Skogsbruk
- Jordbruk
- Gruvindustri
- ...

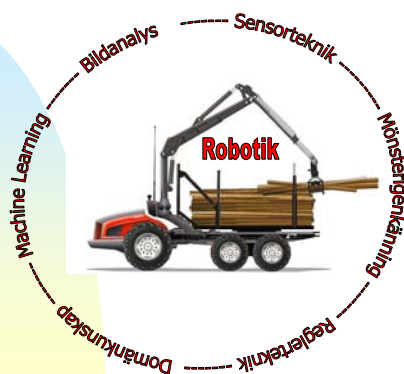
IFOR

Intelligenta Fordon Off-Road

Finansiering:

- 2.0 Mkr Partek Forest
- 3.2 Mkr Umeå University (2 ass. prof.)
- 4.5 Mkr Kempe foundation
(professor/assoc. prof., PhD student)
- 3.0 Mkr Kempe foundation
(project: Autonomous Navigation)

Autonom navigering för skogsmaskiner

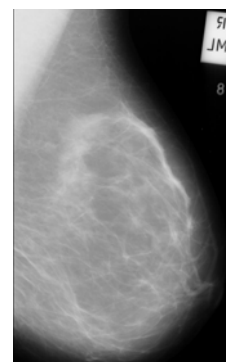
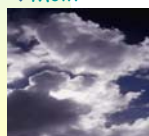


Medicinsk bildanalys och fraktaler

- Fredrik Georgsson, TD, univ. lektor
- Peter Wingren, docent, Matematik
- Vissa naturliga fenomen uppvisar specifika fraktala egenskaper
- Kan vi t.ex. kvantifiera tumörer i röntgenbilder m.h.a. dessa?
- Finansiering: VR'03-05 och tek-nat fakulteten

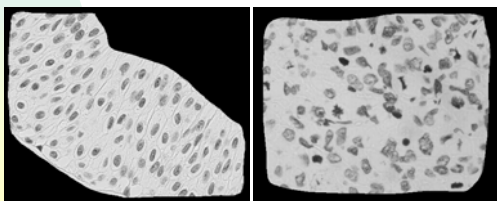
Många naturliga formationer är fraktala

- Exempel
 - ◆ Trädgrenar
 - ◆ Floder
 - ◆ Blodkärl
 - ◆ Kustlinjer
 - ◆ Parenkym
 - ◆ Moln



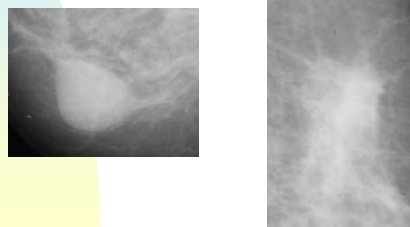
Hypotes 1

- Parenkymet hos en tumör kan uppvisa andra fraktala egenskaper än frisk vävnad



Hypotes 2

- Randen på en massförändring kan uppvisa olika fraktala egenskaper



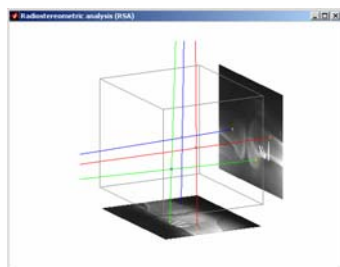
Fraktala egenskaper i bilder - öppna problem

- Naturliga bilder är endast fraktala över vissa skalor, vid autonom beräkning är det inte trivialt att bedöma vilka skalor som är relevanta
- Hur hantera olika intensitet i bilder?
- Olika estimat av fraktala egenskaper borde vara likvärdiga, men ger olika resultat
- Hur projiceras fraktal information? (3D → deformerad 3D → 2D → digital bild)

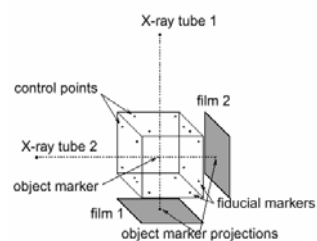
RSA - Stereoröntgen

- Niclas Börnin, FD, univ. lektor
- Per-Åke Wedin, professor
- Röntgenstereofotogrammetri: metod för mätning av mikrorörelser (< 0.1 mm) i skelett
- Prediktera långtidsöverlevnad av implantat
- Modellering: lösning av viktade icke-linjära minstakvadratproblem

Kulor i 3D avbildas på röntgenplåtar

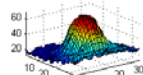


Stereoröntgen med kalibreringsbur

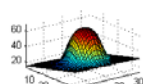


Parameter estimation problems in Radiostereometric Analysis (RSA)

Image data showing a radiostereometric marker



Best model fit. Center of marker is extracted from optimal parameter set.



Teknisk-vetenskapliga och parallella beräkningar

- Mitt forskningsområde ...
- Problem och tillämpningar med stora behov av högpresterande datorresurser
 - ◆ beräkningskapacitet
 - ◆ bearbetning och lagring av stora datamängder
 - ◆ visualisering och VR
- ... i stark utveckling - hand i hand med den tekniska utvecklingen

Skalbara och parallella datorsystem, metoder och programvara ...

- Realistiska modeller av verkligheten konstrueras och analyseras
- Produkter & system utvärderas redan på designstadiet eller då verkliga försök är orealistiska eller olämpliga
- **Forskningsprocessen:** modellering -> datorberäkningar -> visualisering -> analys -> ...
- ... som komplement till teori och experiment!

Forskningsområden

- Numerisk linjär algebra
- Optimering (prof. Per-Åke Wedin)
- Differential-algebraiska system
- Parallella beräkningar
- Ex. tillämpningsområden
 - ◆ Reglerteknik/teori
 - ◆ Stereoröntgen - RSA
 - ◆ Bioinformatik
 - ◆ Realtidssimuleringar (Vista II)
- Växande behov av teori, metoder, algoritmer & programvara

Forskningsgruppen

- Bo Kågström, professor, forsknings-/gruppleddare
- Ji-guang Sun, professor
- Fred Gustavson, adj. prof (IBM T.J. Watson Res. Cen)
- Erik Elmroth, univ. lektor, forskarassistent
- Frederic Gyuomarc'h, visiting post doc
- Björn Adlerborn, forskningsass.
- Robert Granat, doktorand
- Jeanette Hargbo, doktorand (bioinformatik, UCMP)
- Isak Jonsson, TD, vik. lektor
- Pedher Johansson, doktorand
- Stefan Johansson, doktorand
- Claude Lacoursière, doktorand
- Mikael Rännar, FL, adjunkt

Externfinansierade projekt

- Theory, Algorithms, and Tools for Computing Nearby Jordan and Kronecker Structures with Applications in Control Theory, VR: 1.0 MSEK (2001-03/06)
- Hierarchially Blocked Algorithms and Optimized Low-Level Kernels for Dense Matrix Computations on Memory-Tiered High Performance Computing Systems, VR: 1.3 MSEK (2002-04)
- Matrix Pencil Computations in Computer-Aided Control System Design: Theory, Algorithms and Software Tools, SSF ramanslag: 7.5 MSEK (2003-05)
- Generic Grid Computing Research for Resource Management, Portals and Algorithms, VR: 2.0 MSEK (2003-05) **SweGRID**

Parallella beräkningar

... en samling processorer som kommunicerar och samarbetar för att lösa ett stort problem snabbt!

Kan alla problem lösas parallellt?

Gräva ett håll i marken: Ja Nej
Kan Parallelliseras? Ja Nej

Gräva ett dike: Ja Nej
Kan Parallelliseras? Ja Nej

Databeroende: Ja Nej
Kan en sten läggas var som helst när som helst? Ja Nej

Design av parallella program

- Datapartitionering
 - ◆ fördela data på de olika processorerna
- Kornighet
 - ◆ storlek på de oberoende delarna som kan utföras samtidigt
 - ◆ fin- , grovkorning parallellitet
- Lastbalans
 - ◆ se till så att alla processorer har "samma last"

Shared Memory Architectures

- Alla proc. har access till ett globalt minne
 - ◆ UMA, NUMA
- Access till det gemensamma minnet
 - ◆ bus eller switching network
- Hårdvaran skalar inte bra för massiv parallellism

Distributed Memory Architectures

- Varje nod har sitt eget lokala minne
- All kommunikation via meddelanden och paket
- Beräkningar mycket snabbare än kommunikation
- Hårdvaran skalar bra

Våra tjänster på nätet

- Förmedlar information och kunskap
 - ◆ senaste forskningsresultaten
- Problemlösningssmiljöer (PSE)
 - ◆ Du specificerar "vad som ska göras" (ej hur!)
 - ◆ Vi levererar resultatet
 - bästa metod/datorkraft
 - kvantitativa och kvalitativa omdömen

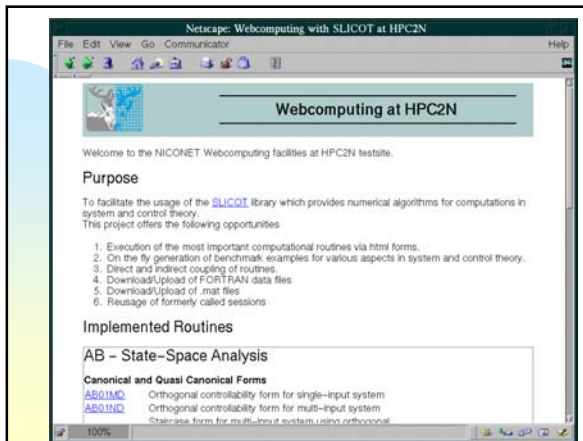
Niconet Webcomputing

- Numerics In COnTrol NETwork
- Tematiskt nätverk, EU
 - ◆ 17 partners (9 universitet, 7 företag)
 - ◆ HPC2N, Umeå universitet
- Utveckla och utvärdera numeriskt tillförlitlig programvara inom regler-teknik och dess implementering i produktionsteknologier

HPC2N-"HPC to North"

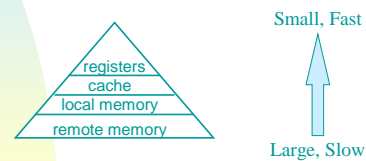


- Nationellt kompetenscentrum för teknisk-vetenskapliga och parallella beräkningar
- Kompetens- & driftsmiljöutveckling
- Tillgodose behov av HPC-resurser
- Forskning och experimentell utveckling
- Visualisering & VR
- Speciellt ansvar för nya områden



Surfar i minneshierarkier ...

- hos dagens datorarkitekturer
 - ◆ PC - kluster - superdatorer



Minimera "frakt" av data i minneshierarkin

- **Arkitekturevolution:** HPC-system med multipla SMP-noder, flera nivåer av cache-minnen, och fler funktionella enheter per CPU
- **Nyckel till prestanda:** förstå interaktionen mellan algoritm och arkitektur
- Biblioteksprogramvara - optimerade byggblock för "basala operationer"

bland andra hierarkiska strukturer

- Egenvärdesproblem
 - ◆ Fiolsträng
 - ◆ En konstruktions hållfasthet (styvhet, elasticitet mm)
- $Ax = sx$ eller $Ax = sBx$
 - ◆ s egenvärde, x egenvektor
- Hierarkier av Jordan- och Kroneckerstrukturer
 - ◆ Orbit-hierarkier


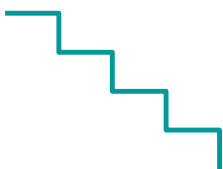
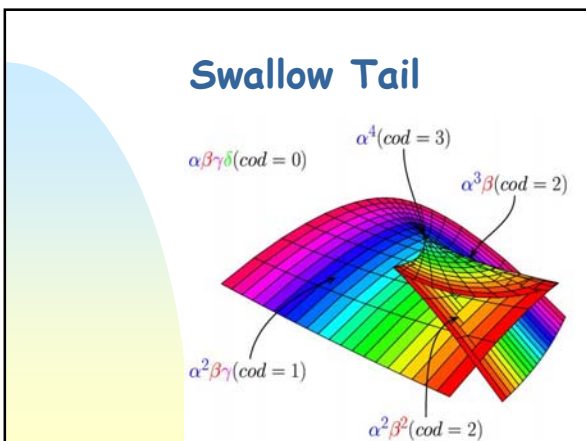
Degeneration via konfluens

- Matrisen A:s kanoniska struktur (JCF) ger information om ett dynamiskt systems uppförande ($y'(t) = Ay(t)$)
- Att beräkna JCF är ett illa ställt problem då vi har flera "lika" egenvärden

$$\begin{bmatrix} \alpha & x & x & x \\ 0 & \alpha & x & x \\ 0 & 0 & \alpha & x \\ 0 & 0 & 0 & \alpha \end{bmatrix} \begin{bmatrix} \alpha & x & x & x \\ 0 & \alpha & x & x \\ 0 & 0 & \beta & x \\ 0 & 0 & 0 & \beta \end{bmatrix} \begin{bmatrix} \alpha & x & x & x \\ 0 & \alpha & x & x \\ 0 & 0 & \beta & x \\ 0 & 0 & 0 & \gamma \end{bmatrix} \begin{bmatrix} \alpha & x & x & x \\ 0 & \alpha & x & x \\ 0 & 0 & \beta & x \\ 0 & 0 & 0 & \gamma \end{bmatrix} \begin{bmatrix} \alpha & x & x & x \\ 0 & \beta & x & x \\ 0 & 0 & \gamma & x \\ 0 & 0 & 0 & \delta \end{bmatrix}$$

Konfluensvarning!

In case of fire, do not use the elevator, ... use the stairs!

Matrisberäkningar och virtuella rum - introduktion

Matrix Computations and Virtual Spaces
 Bo Kågström
 2004-08-30

Matrix
 Matrix Revisited
 Matrix Revolutions
Matrix Computations and Virtual Spaces

Coming soon to a PC near you!

Matrix
 Matrix Revisited
 Matrix Revolutions
Matrix Computations and Virtual Spaces

Coming soon to a PC near you!

Vad är en matris?

1	2	3	4
5	6	7	8

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

1	2
3	4
5	6
7	8

- Matriser av dimension 2×4 , 4×4 och 4×2
- Utdrag från definition på nätet: "A matrix is a two-dimensional array of numbers or more complicated things. There are (easy) rules for adding matrices of the same shape and sometimes (more complicated) rules for multiplying or even dividing them."

Vad är detta?

1780	2200
4116	5080

Några pionjärer inom matrisberäkningar

Från vänster: Jim Wilkinson, Wallace Givens, George Forsythe, Alston Householder, Peter Henrici och Fritz Bauer (1964, Gatlinburg, TN)

Tillämpningar av matrisberäkningar - några smakprov

- Informationssökning (textdokument i digitala bibliotek, sökmotorer för websidor)
- Datorgrafik (2D, 3D)
- Bildbehandling (datakompression, frekvensfiltrering)
- Hållfasthet i bro - konstruktion och analys (Tacoma Narrows Bridge, "egensvängning")
- Positionsnavigering med GPS (optimering)

Informationssökning på internet - världens största matrisberäkning?

Exempel på sökresultat

Hur går en Googlesökning till?

- Google hittar "alla" hemsidor som matchar alla sökorden i en fråga.
- Hemsidorna listas i en ordning som bestäms av PageRank-algoritmen (Larry Page, Sergey Brin - Googles grundare).
- Rankingen bestäms utifrån en $n \times n$ matris A som beskriver länkstrukturen på internet:
 - $A(i,j) = 1$ hyperlänk från sida i till sida j (0 annars)
 - $n > 4.000\ 000\ 000$ (över 4 miljarder)
 - A gles - uppdateras en gång i månaden.
- PageRank är en egenvektor till en Markovprocess ($x = ((1-c)/n E + cA)x$, $c = 0.9$, potensmetoden).

