Parallelisierung Numerischer Verfahren

Volker John

30. März 2001

Inhaltsverzeichnis

Kapitel 1

Motivation

1.1 Rechnertypen und Einsatzgebiete

Es gibt verschiedene Arten von Computern, die sich in ihrer Hardware, ihren Anschaffungskosten und ihren Anwendungsgebieten unterscheiden:

- PC 2000 (1 GHz, 256 MB RAM, 3000 - 5000 DM)
- Workstations 2000 (1 GB RAM, $\approx$ 100000 DM)
- Vektorrechner. Diese Rechner können eine Operation nicht nur mit skalaren Größen durchführen sondern mit ganzen Arrays (Vektoren).

  Beispiel 1.1 Es ist $\mathtt{y} := \mathtt{y} + \alpha\mathtt{x}$ zu berechnen, wobei $\mathtt{x}, \mathtt{y}$ n–dimensionale Arrays und α eine Zahl ist.
  Standardrechner:

  ```
  for(i = 0; i < n; i++)
    y[i] += α * x[i];
  ```

  Vektorrechner mit Registerlänge n.

  Bild

  Die Operationen $*$ und $+$ werden für alle Einträge der Arrays gleichzeitig durchgeführt. Das ist eine Form der Parallelität auf Maschinenebene.

  Vektorrechner verlieren zunehmend an Bedeutung. Sie werden in dieser Vorlesung nicht weiter behandelt.

- **Parallelrechner**
  - **Massiv–Parallelrechner** 2000 (HP N-Class, 8 Prozessoren, 16 GB RAM, 440 MHz, 1760 MFlops peak performance, IZPR: 3 Stück - > 1 Mill. DM)
  - **Cluster von Workstations** (oder PCs)

Aktuelle Übersichten über die Leistungsfähigkeit von Parallelrechnern findet man unter
`http://www.top500.org`
`http://www.specbench.org` (Standard Performance Evaluation Corporation).

Die verschiedenen Rechnertypen besitzen unterschiedliche **Einsatzgebiete**:

- PC : Wirtschaft (Textverarbeitung, Tabellenkalkulation), Schule, Privathaushalte (Spiele)
- Workstations, Vektorrechner, Parallelrechner : Forschung (Universitäten, Wirtschaft), Anwendung der Forschung (Rechenzentren, Wirtschaft, Universitäten), Verwaltung (Lohnabrechnung in der Wirtschaft)

Einige Forschungsgebiete und Anwendungen, die Parallelrechner verwenden sind
- Simulation von Naturprozessen, z. Bsp. **Strömungen**, chemische Reaktionen Anwendungen : Flugzeug– und Kraftfahrzeugdesign (Aerodynamik, Strömungsverhältnisse im Kraftfahrzeug), Ausbreitungen von Umweltverschmutzungen, Wetter– und Klimavorhersage, Computermedizin
- Simulation von industriellen Prozessen, z. Bsp. Schüttguttransport, Brückenkonstruktionen
- Optimierungs– und Suchverfahren, die im wesentlichen auf Probieren (Vergleichen) beruhen
- Bildverarbeitung

Die Simulation von Naturprozessen führt im Kern oft auf die Lösung linearer Gleichungssysteme mit schwach besetzten Matrizen. Die Parallelisierung von Verfahren zur Lösung solcher Gleichungssysteme ist ein Schwerpunkt dieser Vorlesung.

Beispiel 1.2 Navier–Stokes Gleichungen.

1.2 Vorteile, Nachteile und Herausforderungen von Parallelrechnern

Parallelrechner bieten folgende **Vorteile**:
- Mehrere Prozessoren können zur Lösung eines Problems benutzt werden (**Hardwareparallelität**):
  - Problem muß vom Programmierer (oder durch entsprechende Software) in Teilprobleme zerlegt werden,
  - auf jedem Prozessor läuft ein (eigenes) Programm, das eine Anzahl $(0, \dots, n_p)$ Teilprobleme löst,
  - zur Lösung benötigte Daten, die sich auf anderen Prozessoren befinden, müssen sich durch Datenaustausch (Kommunikationen) beschafft werden
- i.a. besitzen Parallelrechner viel mehr Speicher als ein einzelner Rechner $\Longrightarrow$ große Probleme können gelöst werden.

Es gibt jedoch auch einige **Nachteile**. Der Einfluß einiger Nachteile kann durch adequate Auswahl von numerischen Verfahren und gute Softwareentwicklung klein gehalten werden.
- Kommunikationen beeinflussen die Effizienz der verwendeten Verfahren negativ; oft ist Kommunikation langsam im Vergleich zur Rechengeschwindigkeit $\Longrightarrow$ so wenig wie nötig und so geschickt wie möglich kommunizieren !

- vorhandene Software für Einzelrechner muß an den Stellen erweitert werden, an denen Kommunikationsbedarf besteht und neue Software muß unter dem parallelen Blickwinkel konzipiert werden (**algorithmische Parallelität**) :
  - bei der Auswahl der verwendeten Verfahren ist neben deren numerischen Verhalten jetzt auch auf Parallelisierbarkeit und parallele Effizienz zu achten,
  - die ausgewählten Verfahren müssen unter dem Blickwinkel eines minimalen Kommunikationsbedarfes analysiert werden,
  - Kommunikation muß programmiert werden.
- Massiv–Parallelrechner
  - vergleichsweise anfällig für Hard– und Softwarefehler
  - benötigen viel Ausstattung (klimatisierte Räume, hochbelastbare Fußböden, unterbrechungsfreie Stromversorgung, ...),
  - keine einheitlichen Betriebssysteme und vom Hersteller entwickelten Routinen zur Kommunikationssteuerung (aber: es gibt portable Bibliotheken, mit denen man auf den meisten Parallelrechnern die Kommunikation steuern kann, z. Bsp. **MPI**).

Eine Einführung in MPI ist Bestandteil der Vorlesung.

Bevor man ein Programm parallelisiert, muß man sich überlegen, wie die in dem Programm genutzten numerischen Verfahren parallelisiert werden können. Es stellt sich heraus, daß es Verfahren gibt, die sich gut parallelisieren lassen und Verfahren, bei denen eine effiziente Parallelisierung schwierig ist.

Beispiel 1.3 Gut und schlecht parallelisierbare Verfahren. Zu lösen ist das lineare Gleichungssystem

$$Ax = b. \tag{1.1}$$

1. Die LU–Zerlegung (Gauß–Verfahren) mit Spaltenpivotsuche zur Lösung von (1.1) ist für kleine Gleichungssysteme (mit bis zu mehreren Hundert Unbekannten) ein numerisch stabiles und effizientes Verfahren. Dieses Verfahren läßt sich jedoch nur schwierig und mit hohen Kommunikationsverlusten parallelisieren.
2. Das Jacobi–Verfahren zur Lösung von (1.1)

   $$x^{n+1} = x^n - \operatorname{diag}(A)^{-1}(Ax^n - b).$$

   ist ein Iterationsverfahren, das falls es konvergiert, sehr langsam konvergiert. Es läßt sich aber einfach und mit wenig Kommunikationsverlusten parallelisieren.

Die im Mittelpunkt der Vorlesung stehenden großen Gleichungssysteme mit schwach besetzten Matrizen werden im allgemeinen iterativ gelöst. Die verwendeten Iterationsverfahren lassen sich relativ gut parallelisieren.