Reducing Memory Access Latencies using Data Compression in Sparse, Iterative Linear Solvers

An All-College Thesis

College of Saint Benedict/Saint John's University

by Neil Lindquist
April 2018

Project Title: Reducing Memory Access Latencies using Data Compression in Sparse, Iterative Linear Solvers
Approved by:

_____

Mike Heroux
Scientist in Residence

_____

Robert Hesse
Associate Professor of Math

_____

Jeremy Iverson
Assistant Professor of Computer Science

_____

Bret Benesh
Chair, Department of Mathematics

_____

Imad Rahal
Chair, Department of Computer Science

_____

Director, All College Thesis Program

Abstract

Contents

1 Introduction

2 Background

2.1 Conjugate Gradient

Conjugate Gradient is the iterative solver used by HPCG [1]. Symmetric, positive definite matrices will guarantee the converge of Conjugate Gradient to the correct solution within n iterations when using exact algebra [3]. As an iterative method, Conjugate Gradient can provide a solution, $\vec{x}$, where $\left\| \mathbf{A}\vec{x} - \vec{b} \right\|$ is within some tolerance, after significantly fewer than n iterations, allowing it to find solutions to problems where even n iterations is infeasible [4].

To understand the Conjugate Gradient, first consider the quadratic form of $\mathbf{A}\vec{x} = \vec{b}$. The quadratic form is a function $f : \mathbb{R}^n \to \mathbb{R}$ where

$$f(\vec{x}) = \frac{1}{2}\vec{x} \cdot (\mathbf{A}\vec{x}) - \vec{b} \cdot \vec{x} \tag{1}$$

for some $c \in \mathbb{R}$. Note that

$$\nabla f(\vec{x}) = \nabla \left(\frac{1}{2}\vec{x} \cdot (\mathbf{A}\vec{x}) - \vec{b} \cdot \vec{x} \right)$$

$$= \frac{1}{2} \left(\mathbf{A}\vec{x} + \mathbf{A}^T \vec{x} \right) - \vec{b}$$

Then, when $\mathbf{A}$ is symmetric,

$$\nabla f(\vec{x}) = \mathbf{A}\vec{x} - \vec{b}$$

So, the solution to $\mathbf{A}\vec{x} = \vec{b}$ is the sole critical point of f [2]. Since $\mathbf{A}$ is the Hessian matrix of f at the point, if $\mathbf{A}$ is positive definite, then that critical point is a minimum. Thus, if $\mathbf{A}$ is a symmetric, positive definite matrix, then the minimum of f is the solution to $\mathbf{A}\vec{x} = \vec{b}$ [4].

To understand Conjugate Gradient, start by considering the Method of Steepest Decent. This method takes an initial $\vec{x}_0$ and repeatedly computes improved points, $\vec{x}_1, \vec{x}_2, \ldots$, until reaching a point close enough to the minimum of Equation 1 [2]. Because the gradient at a point is the direction of maximal increase,

$$\vec{x}_{i+1} = \vec{x}_i + \alpha_i \vec{r}_i \tag{2}$$

for some $\alpha_i > 0$ and where $\vec{r}_i = -\nabla f(\vec{x}_i) = \vec{b} - \mathbf{A}\vec{x}_i$ is the residual of $\vec{x}_i$. So, the value for α_i that minimizes $f(\vec{x}_{i+1})$ occurs at the bottom of the parabola

formed when taking $\vec{x}_{i+1}$ as a function of α_i. Thus, choose α_i such that

$$
\begin{aligned}
0 &= \frac{\mathrm{d}}{\mathrm{d}\alpha_i} f\left(\vec{x}_{i+1}\right) \\
&= \frac{\mathrm{d}}{\mathrm{d}\alpha_i} f\left(\vec{x}_i + \alpha \vec{r}_i\right) \\
&= \nabla f\left(\vec{x}_i + \alpha_i \vec{r}_i\right) \cdot \vec{r}_i \\
&= \left(\mathbf{A}\vec{x}_i - \vec{b}\right) \cdot \vec{r}_i + \alpha_i \mathbf{A}\vec{r}_i \cdot \vec{r}_i \\
\alpha_i &= \frac{\vec{r}_i \cdot \vec{r}_i}{\vec{r}_i \cdot \mathbf{A}\vec{r}_i}.
\end{aligned}
$$

Note that by using Equation 2, we can derive

$$
\vec{r}_{i+1} = \vec{r}_i - \alpha \mathbf{A}\vec{r}_i.
$$

Because $\mathbf{A}\vec{r}_i$ is already computed to find α_i, one less matrix-vector product is needed per iteration. The resulting steps for the Method of Steepest Decent are

$$
\begin{aligned}
\vec{r}_0 &= \vec{b} - \mathbf{A}\vec{x}_0 \\
\alpha_i &= \frac{\vec{r}_i \cdot \vec{r}_i}{\vec{r}_i \cdot \mathbf{A}\vec{r}_i} \\
\vec{x}_{i+1} &= \vec{x}_i + \alpha_i \vec{r}_i \\
\vec{r}_{i+1} &= \vec{r}_i - \alpha \mathbf{A}\vec{r}_i
\end{aligned}
$$

until $\|\vec{r}_i\|$ is less than some tolerance [4].

Example 1. Consider the linear system

$$
\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} 5 \\ 5 \end{bmatrix}
$$

and use $c = 0$. Note that the solution is

$$
\vec{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.
$$

When starting at the origin, the iteration of Method of Steepest Decent becomes

$$
\begin{array}{lll}
\vec{x}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} & \vec{r}_0 = \begin{bmatrix} 5 \\ 5 \end{bmatrix} & \alpha_0 = 2/7 \\[2mm]
\vec{x}_1 = \begin{bmatrix} 10/7 \\ 10/7 \end{bmatrix} & \vec{r}_1 = \begin{bmatrix} 5/7 \\ -5/7 \end{bmatrix} & \alpha_1 = 2/3 \\[2mm]
\vec{x}_2 = \begin{bmatrix} 40/21 \\ 20/21 \end{bmatrix} & \vec{r}_2 = \begin{bmatrix} 5/21 \\ 5/21 \end{bmatrix} & \alpha_2 = 2/7 \\[2mm]
\vec{x}_3 = \begin{bmatrix} 290/147 \\ 50/49 \end{bmatrix} & \vec{r}_3 = \begin{bmatrix} 5/147 \\ -5/147 \end{bmatrix} & \alpha_3 = 2/3 \\[2mm]
\quad\vdots & \quad\vdots & \quad\vdots
\end{array}
$$

The $\vec{x}_i$'s are plotted with a contour graph of the quadratic form in Figure 1.
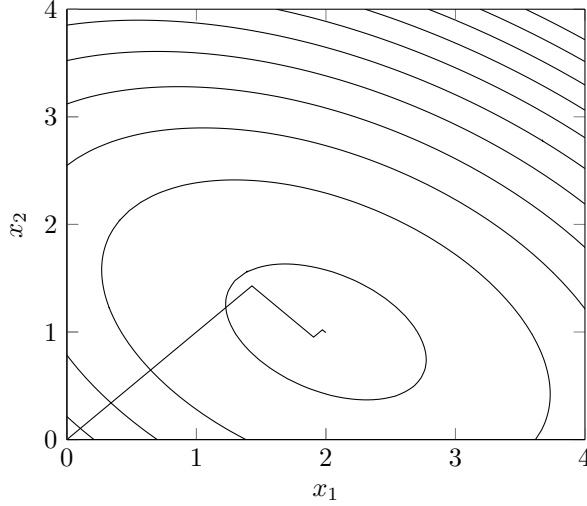
5

Figure 1: Contour graph of the quadratic function and the first six values of $\vec{x}$ produced by steepest decent for Example 1

Next, consider Conjugate Directions, a class of iterative linear solvers, of which Conjugate Gradient is a member of [4]. Note that, in Example 1, the directions of $\vec{r}_0$ and $\vec{r}_2$ are the same and the directions of $\vec{r}_1$ and $\vec{r}_3$ are the same. Thus, the same direction has to be traversed multiple times. Additionally, note that the two sets of residual directions are perpendicular to each other. Conjugate Directions attempts to improve on this, by making the search directions, $\vec{d}_0, \vec{d}_1, \ldots$, $\mathbf{A}$-orthogonal to each other and only moving $\vec{x}$ once in each search direction. Two vectors, $\vec{u}, \vec{v}$ are $\mathbf{A}$-orthogonal, or conjugate, if $\vec{u}^T \mathbf{A} \vec{v} = 0$. The requirement for Conjugate Directions is to make $\vec{e}_{i+1}$ $\mathbf{A}$-orthogonal to $\vec{d}_i$, where $\vec{e}_i = \vec{x}_i - \mathbf{A}^{-1} \vec{b}$ is the error of the current $\vec{x}$. Then, we can derive

$$\alpha_i = \frac{\vec{d}_i^T \vec{r}_i}{\vec{d}_i^T \mathbf{A} \vec{d}_i}.$$

Conjugate Gradient is a form of Conjugate Directions where the residuals are made $\mathbf{A}$-orthogonal to each other. The conjugate Gram-Schmidt process is used to make the residuals $\mathbf{A}$-orthogonal. To do this, each search direction, $\vec{d}_i$ is computed by taking $\vec{r}_i$ and removing any components that are not $\mathbf{A}$-orthogonal to the previous $\vec{d}$'s. So, let $\vec{d}_0 = \vec{r}_0$ and for $i > 0$ let

$$\vec{d}_i = \vec{r}_i + \sum_{k=0}^{i-1} \beta_{ik} \vec{d}_k$$

6

with β_{ik} defined for $i > k$. Then, β_{ik} can be solved for, giving

$$\beta_{ik} = -\frac{\vec{r}_i^T \mathbf{A} \vec{d}_i}{\vec{d}_j^T \mathbf{A} \vec{d}_j}.$$

Note that each residual is orthogonal to the previous search directions and thus previous residuals. So, it can be shown that $\vec{r}_{i+1}$ is $\mathbf{A}$-orthogonal to all previous search directions, except $\vec{d}_i$ [4]. Then, $\beta_{ik} = 0$ for $i \neq k$. To simplify notation, let $\beta_i = \beta_{i,i-1}$. So, each new search direction can then be computed by

$$\vec{d}_i = \vec{r}_i + \beta_i \vec{d}_{i-1}.$$

This results in the following algorithm for Conjugate Gradient

$$\vec{d}_0 = \vec{r}_0 = \vec{b} - \mathbf{A}\vec{x}_0$$
$$\alpha_i = \frac{\vec{r}_i \cdot \vec{r}_i}{\vec{d}_i \cdot \mathbf{A}\vec{d}_i}$$
$$\vec{x}_{i+1} = \vec{x}_i + \alpha_i \vec{d}_i$$
$$\vec{r}_{i+1} = \vec{r}_i - \alpha \mathbf{A}\vec{d}_i$$
$$\beta_{i+1} = \frac{\vec{r}_{i+1} \cdot \vec{r}_{i+1}}{\vec{r}_i \cdot \vec{r}_i}$$
$$\vec{d}_{i+1} = \vec{r}_{i+1} + \beta_{i+1}\vec{d}_i,$$

iterating until $\|\vec{r}_i\|$ is less than some tolerance.

Example 2. Solving the problem from Example 1 using Conjugate Gradient results in

$$\vec{x}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \vec{r}_0 = \begin{bmatrix} 5 \\ 5 \end{bmatrix} \qquad\qquad \vec{d}_0 = \begin{bmatrix} 5 \\ 5 \end{bmatrix} \qquad \alpha_0 = 2/7$$

$$\vec{x}_1 = \begin{bmatrix} 10/7 \\ 10/7 \end{bmatrix} \quad \vec{r}_1 = \begin{bmatrix} 5/7 \\ -5/7 \end{bmatrix} \quad \beta_1 = 1/49 \quad \vec{d}_1 = \begin{bmatrix} 40/49 \\ -30/49 \end{bmatrix} \quad \alpha_1 = 7/10$$

$$\vec{x}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \qquad \vec{r}_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Note that after two iterations, $\vec{x}$ reaches the exact solution, compared to the iteration of Steepest Decent. The iterations are plotted in Figure 2

3 Test Results

4 Conclusions and Future Work

5 References

[1] Jack Dongarra, Michael Heroux, and Piotr Luszczek. Hpcg benchmark: a new metric for ranking high performance computing systems. Technical
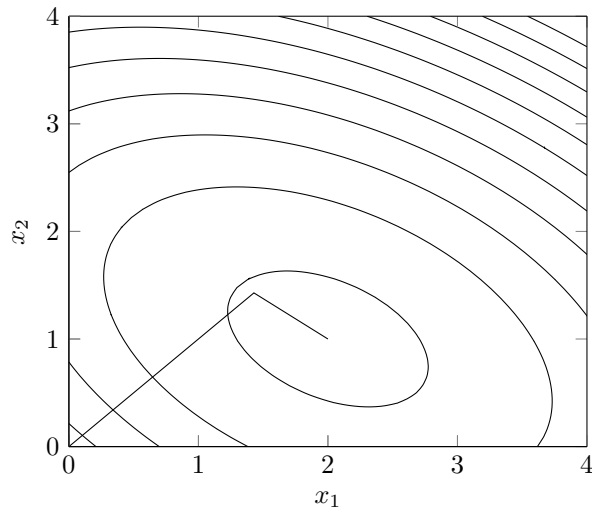
Figure 2: Contour graph of the quadratic function and the each value of $\vec{x}$ produced by Conjugate Gradient for Example 2

Report UT-EECS-15-736, Electrical Engineering and Computer Sciente Department, Knoxville, Tennessee, November 2015.

[2] J. Nearing. *Mathematical Tools for Physics*. Dover books on mathematics. Dover Publications, 2010.

[3] Y. Saad. *Iterative Methods for Sparse Linear Systems*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2nd edition, 2003.

[4] Jonathan R Shewchuk. An introduction to the conjugate gradient method without the agonizing pain. Technical report, Carnegie Mellon University, Pittsburgh, PA, USA, 1994.