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, where n is the number of dimensions, when using exact algebra [3]. More importantly, Conjugate Gradient can be used as in iterative method, providing 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

$$f(\vec{x})=\tfrac{1}{2}\vec{x}^T\mathbf{A}\vec{x}-\vec{b}\cdot\vec{x}+c \tag{1}$$

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

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

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].

The method of Steepest Decent is useful for understanding Conjugate Gradient, because they both use a similar approach to minimize Equation 1, and thus solve $\mathbf{A}\vec{x}=\vec{b}$. This shared approach is to take an initial $\vec{x}_0$ and move downwards in the steepest direction, within certain constraints, of the surface defined by Equation 1 [2]. Because the gradient at a point is the direction of maximal increase, $\vec{x}$ should be moved in the opposite direction of the gradient. Thus, to compute the next value of $\vec{x}$, use

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

for some $\alpha_i>0$ and where $\vec{r}_i=-\nabla f\left(\vec{x}_i\right)=\vec{b}-\mathbf{A}\vec{x}_i$ is the residual of $\vec{x}_i$. Since $\mathbf{A}\vec{x}=\vec{b}$ is the only critical point and a minimum of the quadratic function, f, the ideal value of α_i is the one that minimizes $f\left(\vec{x}_{i+1}\right)$. Thus, choose α_i such that

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

Note that by using Equation 2, we can derive

$$\vec{r}_{i+1} = \vec{r}_i - \alpha \mathbf{A} \vec{r}_i. \tag{3}$$

Because $\mathbf{A}\vec{r}_i$ is already computed to find α_i, using Equation 3 to compute the residual results in one less matrix-vector product per iteration. The 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

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

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

The Conjugate Directions family of linear solvers, of which Conjugate Gradient is a member of, attempts to improve on the number of iterations needed by Steepest Decent. [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

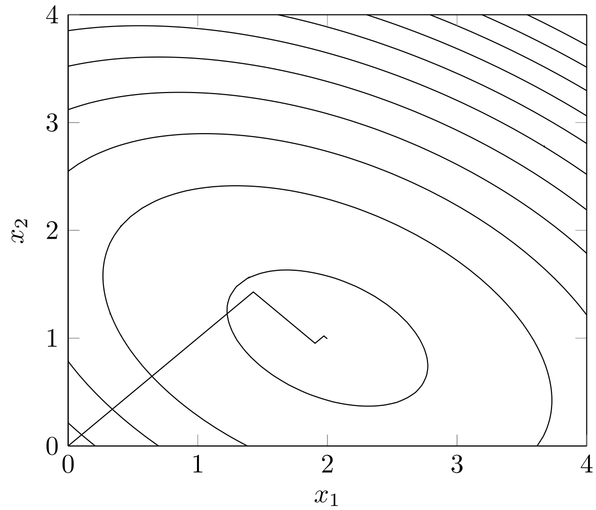

Figure 1: Contour graph of the quadratic function and the first six values of $\vec{x}$ produced by steepest decent for Example 1

to improve on this, by making the search directions, $\vec{d}_0, \vec{d}_1, \ldots,$ **A**-orthogonal to each other and only moving $\vec{x}$ once in each search direction. Two vectors, $\vec{u}, \vec{v}$ are **A**-orthogonal, or conjugate, if $\vec{u}^T \mathbf{A} \vec{v} = 0$. The requirement for Conjugate Directions is to make $\vec{e}_{i+1}$ **A**-orthogonal to $\vec{d}_i$, where $\vec{e}_i = \vec{x}_i - \mathbf{A}^{-1}\vec{b}$ is the error of $\vec{x}_i$. The computation of α_i changes to find the minimal value along $\vec{d}_i$ instead of $\vec{r}_i$.

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

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

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

with $\beta_{(i,k)}$ defined for $i > k$. Then, solving for $\beta_{(i,k)}$ gives

$$\beta_{(i,k)} = -\frac{\vec{r}_i \cdot \mathbf{A}\vec{d}_i}{\vec{d}_j \cdot \mathbf{A}\vec{d}_j}.$$

Note that each residual is orthogonal to the previous search directions, and thus the previous residuals. So, it can be shown that $\vec{r}_{i+1}$ is **A**-orthogonal to all previous search directions, except $\vec{d}_i$ [4]. Then, $\beta_{(i,k)} = 0$ for $i - 1 \neq k$. 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 steps for Conjugate Gradient

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

continuing the iteration until $\|\vec{r}_i\|$ is less than the specified tolerance.

Example 2. Consider the linear system used in Example 1 where

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

The result of applying Conjugate Gradient is

$$\begin{array}{lllll}
\vec{x}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} & \vec{r}_0 = \begin{bmatrix} 5 \\ 5 \end{bmatrix} & & \vec{d}_0 = \begin{bmatrix} 5 \\ 5 \end{bmatrix} & \alpha_0 = 2/7 \\
\vec{x}_1 = \begin{bmatrix} 10/7 \\ 10/7 \end{bmatrix} & \vec{r}_1 = \begin{bmatrix} 5/7 \\ -5/7 \end{bmatrix} & \beta_1 = 1/49 & \vec{d}_1 = \begin{bmatrix} 40/49 \\ -30/49 \end{bmatrix} & \alpha_1 = 7/10 \\
\vec{x}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} & \vec{r}_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} & & &
\end{array}$$

Note that after two iterations, $\vec{x}$ reaches the exact solution, compared to the iterations of Steepest Decent in Example 1. Figure 2 shows the values of $\vec{x}$ with the contour graph of the quadratic function.

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

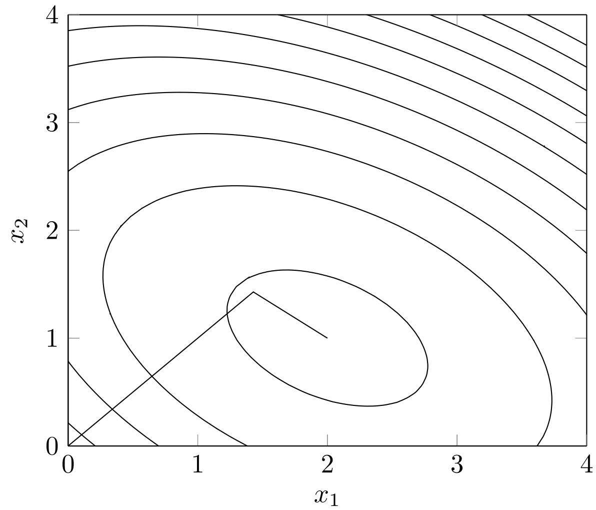

Figure 2: Contour graph of the quadratic function and the each value of $\vec{x}$ produced by Conjugate Gradient for Example 2

[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.