Adaptation of the NVIDIA CUDA Engine to the Single-Step MaxT Multiple Hypothesis Testing Procedure

by

William L. Welbourn, Jr.

Department of Mathematics and Statistics
Utah State University
Logan, Utah
bill.welbourn@aggiemail.usu.edu

May 5, 2010

Abstract

A genome wide association study (GWAS), in which the investigator seeks to determine genetic markers associated with a disease trait, can be restated as a problem in multiple hypothesis testing (MHT): the simultaneous test for each genetic marker of the null hypothesis of no association between the marker and disease. When many hypotheses are simultaneously tested, where each test has a specified Type I error probability, the chance of committing some Type I errors increases. A GWAS can consist of testing hundreds of thousands of genetic markers, thereby presenting a large multiplicity problem.

MHT procedures have been developed to control for Type I error when multiple hypotheses are simultaneously tested. Common MHT procedures to control the family-wise Type I error rate include the Bonferroni, Šidák, and minP and maxT permutation procedures. Due to their extraordinary computational burden within a GWAS, it is tempting to abandon permutation MHT procedures in lieu of an easily adaptable MHT procedure, such as the Bonferroni procedure. As a consequence in doing this, one can incur a loss of statistical power.

This article discusses the development of an algorithm for adapting the maxT MHT procedure to the NVIDIA[1] CUDA programming environment. The algorithm is applied to five (5) statistical tests accessible within the publicly available PLINK[2][Purcell et al. (2007)] software package, against a case-control GWAS data set comprised of 697,226 single nucleotide polymorphism (SNPs) markers and 1,001 cases and 1,034 controls. At a minimum, the results indicate a realized speedup of more than 125 times, utilizing the CUDA algorithm over PLINK.

[1]Source: http://www.nvidia.com/cuda; implementation to a single NVIDIA GeForce GTX 470 graphics card
[2]Source: http://pngu.mgh.harvard.edu/~purcell/plink.