Int J Comput Vis
DOI 10.1007/s11263-013-0643-y

Sampling Minimal Subsets with Large Spans for Robust Estimation

Quoc Huy Tran · Tat-Jun Chin · Wojciech Chojnacki · David Suter

Received: 9 June 2012 / Accepted: 10 July 2013

Abstract When sampling minimal subsets for robust parameter estimation, it is commonly known that obtaining an all-inlier minimal subset is not sufficient; the points therein should also have a large spatial extent. This paper investigates a theoretical basis behind this principle, based on a little known result which expresses the least squares regression as a weighted linear combination of all possible minimal subset estimates. It turns out that the weight of a minimal subset estimate is directly related to the span of the associated points. We then derive an analogous result for *total least squares* which, unlike ordinary least squares, corrects for errors in both dependent and independent variables. We establish the relevance of our result to computer vision by relating total least squares to geometric estimation techniques. As practical contributions, we elaborate why naive distance-based sampling fails as a strategy to maximise the span of all-inlier minimal subsets produced. In addition we propose a novel method which, unlike previous methods, can consciously target all-inlier minimal subsets with large spans.

Keywords Least squares · Total least squares · Minimal subsets · Robust fitting · Hypothesis sampling

Q. H. Tran · T.-J. Chin (✉) · W. Chojnacki · D. Suter
The Australian Centre for Visual Technologies,
and School of Computer Science, The University of Adelaide,
North Terrace, SA 5005, Australia
e-mail: tatjun@gmail.com; tjchin@cs.adelaide.edu.au

Q. H. Tran
e-mail: huy@cs.adelaide.edu.au

W. Chojnacki
e-mail: wojtek@cs.adelaide.edu.au

D. Suter
e-mail: dsuter@cs.adelaide.edu.au

1 Introduction

One of the earliest recorded usage of minimal subsets in statistical estimation occurred in 1755, when Boscovich attempted to determine the meridian arc near Rome from five measurements (Stigler 2000). He solved for the two unknowns of the arc using all ten possible pairings of the data. Two of the pairs were ignored for yielding what Boscovich considered to be unusual outcomes, and the remaining estimates were simply averaged for his final result. Boscovich's work predated Gauss's paper on least squares (published 1809), but did not gain traction due to a lack of analytical basis.

Presently however, the usage of minimal subsets has become an integral part of robust parameter estimation, especially in computer vision for the estimation of multiple view geometry from noisy images (Hartley and Zisserman 2004). This stems from the fact that many robust criteria (e.g., least median squares Rousseeuw and Leroy 1987, maximum consensus Fischler and Bolles 1981) do not have closed form solutions. Also, many geometric models of interest (e.g., fundamental matrix) have a large number of parameters, thus sampling and testing model hypotheses from minimal subsets is often the only way to obtain good solutions in reasonable time.

Intuitively, drawing an *all-inlier* minimal subset is not sufficient to guarantee a reasonably good model hypothesis; the inliers therein should also have a large span. To illustrate this notion, consider the problem of line fitting on the 2D data in Fig. 1, where the data has been generated without outliers for simplicity. Two particular choices of (all-inlier) minimal subsets are highlighted; clearly Set A yields a better estimate than Set B, as can be verified by a suitable goodness-of-fit function (e.g., Fischler and Bolles 1981; Rousseeuw and Leroy 1987). It is also

Published online: 27 July 2013