FP7-PEOPLE-2009-IRSES:
Project ID 246647

Optimization and its Applications in Learning and Industry (OptALI)

IRSES

Ongoing Deliverable D1.2

Description of Research Seminar: Line location with outliers

Start date of the Workpackage:	December 2010
Duration:	48 months
Due date of deliverable:	November 2014
Actual submission date:	April 2013
Participants:	UGOE UNIKL DTU UOA UC
Author of deliverable:	R. Schieweck (r.schieweck@math.uni-goettingen.de)

Research Seminar

offered by Robert Schieweck (ES-UGOE-6)

in February 2013,

in Christchurch, New Zealand

Subject: Line location with outliers

Problem: In statistical estimation processes the observed data points may sometimes be subject to blatant measurement errors which cannot be modeled via the usual approach that all observations originate from a certain distribution whose parameters need to be estimated. Such spurious data points have to be treated separately from the bulk of observations and identified as so called outliers in the estimation process. In the case of linear regression in the plane this is often done by iteratively reweighted least squares estimation which down-weights possible outliers. We consider a new approach which is not based on the least squares paradigm but instead combines the the concept of trimming – which is the process of identifying outliers and removing them from the computation of the regression line – and the usage of estimators suited for long-tailed error distributions. Thus our estimators are robust in a double sense. Their computation amounts to solving a location problem where a line has to be located in the plane in order to minimize the sum of distances to a subset of fixed size of some finite given point set.

Main Results: While a straight-forward adaption of the techniques for locating a line without taking into account the presence of outliers only results in an exponential running time, a closer inspection yields polynomial algorithms of low degree. These rely particularly on the *weak incidence property* and the *pseudo-halving property* which are well-known for line location problems without outliers. While the *weak incidence property*

directly carries over to the trimming case an yields a $O(n^3)$ algorithm for the general problem including outliers via a finite candidate set which can be enumerated, the *pseudo-halving property* is of particular use in the unweighted version of the problem where a time reduction to $O(n^2 \log n)$ can be achieved. Moreover, a finite candidate set can also be derived for the problem of location multiple lines while still incorporating the trimming process. Even though the resulting algorithms are of exponential running time – which is expected since the problem is NP hard – they are the first exact algorithms for the problem so far.

Participants: researchers from UC

Publication: -