Efficient robust digital hyperplane fitting with bounded error

Dror Aiger, Yukiko Kenmochi, Lilian Buzer, Hugues Talbot
Université Paris-Est, Laboratoire d'Informatique Gaspard-Monge, France

Contribution

Given a set of N points in a digital image containing noise, we consider the problem of **digital hyperplane fitting**. We first observe that

- it probably cannot be solved **exactly** with the computational complexity better than $O(N^d)$ in d dimension,

and therefore we propose

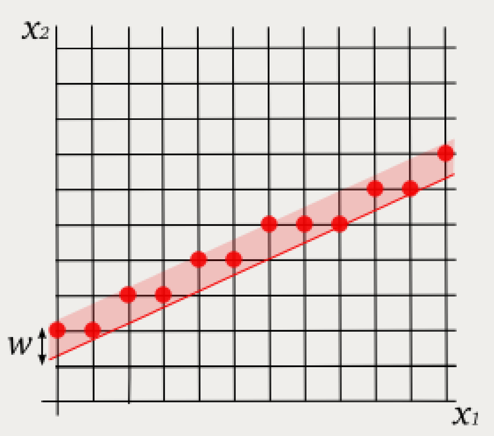
- **two approximation methods** featuring linear time complexity with different bounded errors.

Digital hyperplane

A **digital hyperplane** $\mathbf{D}(\mathbf{H})$ is defined by the set of discrete points satisfying two linear inequalities:

$$\mathbf{D}(\mathbf{H}) = \{(x_1, x_2, \ldots, x_d) \in \mathbb{Z}^d : 0 \le a_1x_1 + a_2x_2 + \ldots + a_dx_d + a_{d+1} < w\} \quad (1)$$

with the normalization $-1 \le a_i \le 1$ for $i = 1, 2 \ldots, d$ such that there exists at least one coefficient $a_i = 1$, where w is a given constant (we set $w = 1$ from digital geometrical viewpoint).



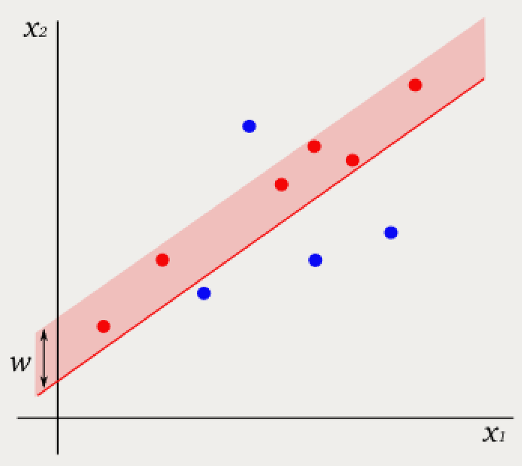
An example of digital lines.

$\mathbf{D}(\mathbf{H})$ is also interpreted as a **w-slab**, which is defined as the region on and in between two parallel hyperplanes of distance w.

Fitting problem

in the primal space

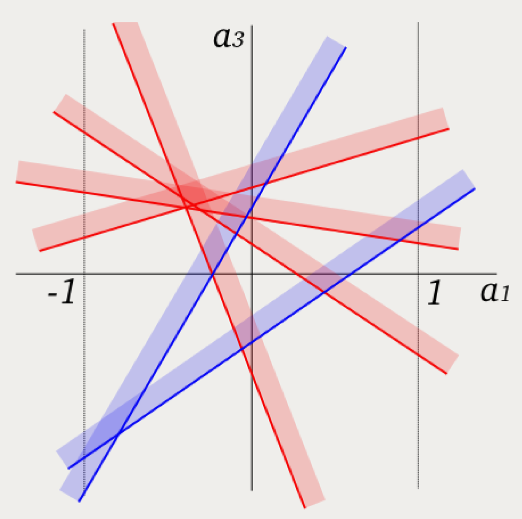
Given a set $\mathbf{S}$ of discrete points coming from the $[0, \delta]^d$ grid, we seek a $\mathbf{D}(\mathbf{H})$ or w-slab that contains a maximum number of points in $\mathbf{S}$.



Points $\boldsymbol{p} \in \mathbf{S}$ are called **inliers** if $\boldsymbol{p} \in \mathbf{D}(\mathbf{H})$. Our problem is then equivalent to finding a $\mathbf{D}(\mathbf{H})$ such that the number of inliers be maximum, called an **optimal digital hyperplane**.

in the dual space

Using geometric duality induced by (1), the problem in the dual space is to find a point that is covered by a maximum number of w-slabs (distance w in the a_{d+1}-axis direction).



Theoretical observation on exact fitting

There is the computational problem called 3SUM, which is conjectured to require roughly quadratic time complexity, and a problem is called **3SUM-hard** if solving it in subquadratic time implies a subquadratic time algorithm for 3SUM.

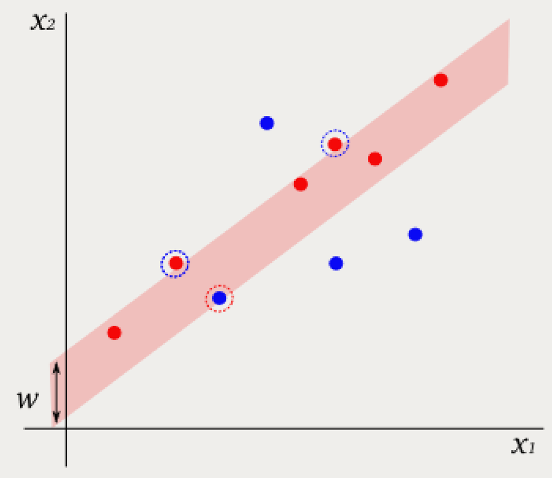
Observation: *The problem of digital line fitting is 3SUM-hard.*

The extension to higher dimension d can be also considered; the **exact solution** of the digital hyperplane fitting problem is likely as hard to be obtained as that of $O(N^d)$ problems.

Approximation with bounded error in number of inlier points

We reduce the fitting problem in the dual space to the problem of **linear programming with violations**, and solve it using the result obtained by Aronov et al.(2008).

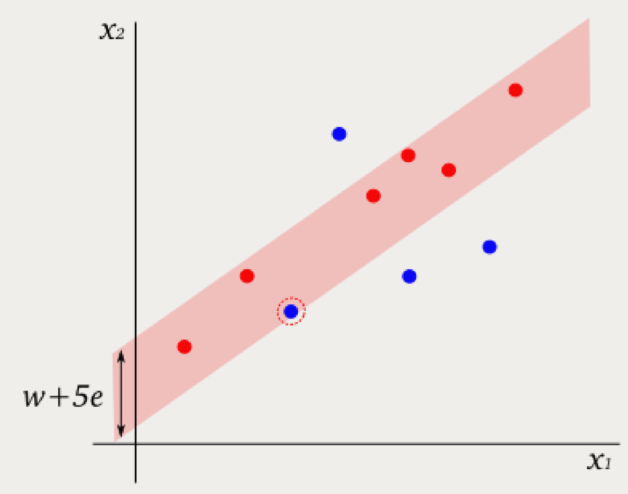
Observation: *Given a set of N points in the d-dimensional space and some $\varepsilon > 0$, we can find a **w-slab that contains at least** $(1-\varepsilon)n_{opt}$ **points**, where n_{opt} is the maximum possible number of points that can be found in a w-slab, assuming $n_{opt} = \Omega(N)$. The runtime is $O(N + \varepsilon^{-4}\log N)$ for $d = 2$ and $O(N(\varepsilon^{-2}\log N)^{d+1})$ for larger d.*



Approximation with bounded error in digital hyperplane width w

We consider the fitting problem as range searching and make use of the data structure introduced by Fonseca and Mount (2010) for their approximate method.

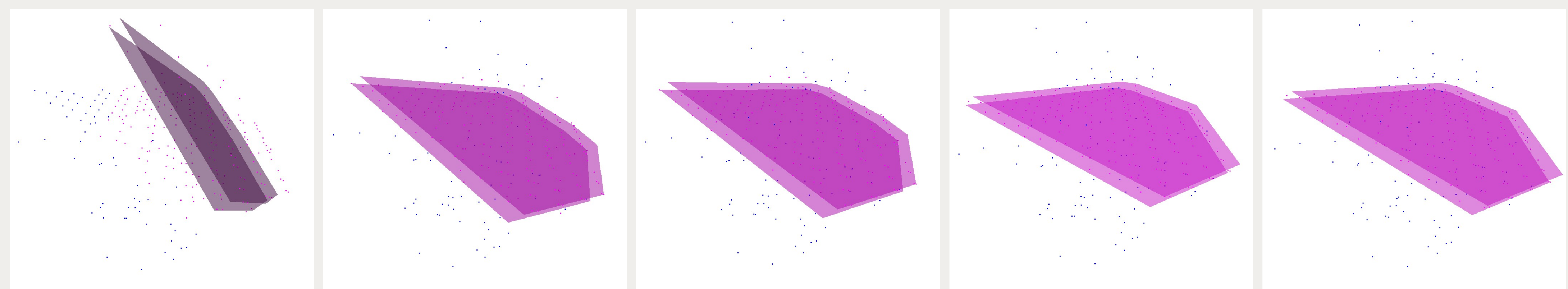
Theorem: *Given a set of N points on a grid $[0, \delta]^d$, and some $\varepsilon > 0$, $w > 0$, a **digital hyperplane of width** $w + 5\varepsilon$ **that contains** $n > n_{opt}$ **points**, can be found in $O(N + (\frac{\delta}{\varepsilon})^d \log^{O(1)}(\frac{\delta}{\varepsilon}))$ time where n_{opt} is the maximum number of points that any digital hyperplane of width w in $[0, \delta]^d$ can contain.*



Typically, we use $0 < \varepsilon < 0.5$ for $w = 1$.

Experiments with the second approximation method

Experimental results of digital plane fitting for a 3D synthetic volume data generated by 400 points in a digital plane and 100 randomly generated points.

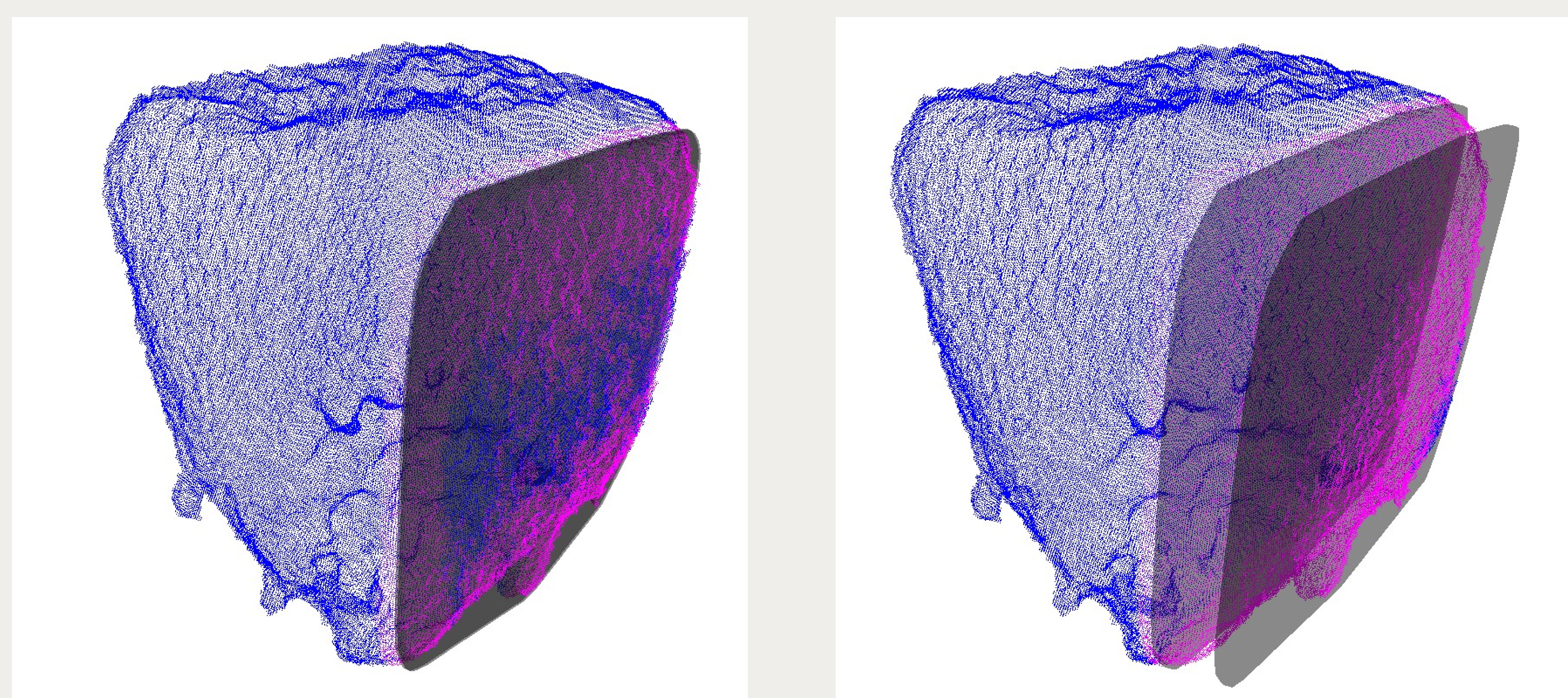


$\varepsilon = 2.0$ | $\varepsilon = 1.0$ | $\varepsilon = 0.5$ | $\varepsilon = 0.25$ | Exact solution

The runtimes, parameters and numbers of points of fitted digital planes.

	runtime	parameters a_1	a_2	a_3	a_4	nb. of points w	$w+5\varepsilon$
Approximate fitting							
$\varepsilon = 2.0$	31 msec	-0.0625	-0.0625	1	-10.9	110	485
$\varepsilon = 1.0$	266 msec	-0.5	-0.5	1	-3.0002	300	435
$\varepsilon = 0.5$	2172 msec	-0.5	-0.5	1	-3.0002	300	421
$\varepsilon = 0.25$	16640 msec	-0.5625	-0.546875	1	-2.55002	362	410
Exact fitting							
with exact comp.	35 min 29.109 sec	$-47/81$	$-43/81$	1	$-203081/81000$	406	
without exact comp.	4 min 36.908 sec	-0.580247	-0.530864	1	-2.507173		

Results of digital plane fitting for a pre-processed 3D binary nano-tomography image of $512 \times 511 \times 412$ containing 205001 discrete points.



$\varepsilon = 4$ for $w = 1$ (left) and $w = 25$ (right).