



HAL
open science

Solving the Geometric Docking Problem for Planar and Spatial Sets

Michel Petitjean

► **To cite this version:**

Michel Petitjean. Solving the Geometric Docking Problem for Planar and Spatial Sets. Internet Electronic Journal of Molecular Design, 2002, 1 (4), pp.185 - 192. hal-02115007

HAL Id: hal-02115007

<https://hal.science/hal-02115007>

Submitted on 15 May 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

Internet Electronic Journal of Molecular Design

April 2002, Volume 1, Number 4, Pages 185–192

Editor: Ovidiu Ivanciuc

Special issue dedicated to Professor Alexandru T. Balaban on the occasion of the 70th birthday
Part 4

Guest Editor: Mircea V. Diudea

Solving the Geometric Docking Problem for Planar and Spatial Sets

Michel Petitjean

ITODYS (CNRS, ESA 7086), 1 rue Guy de la Brosse, 75005 Paris, France

Received: October 16, 2001; Accepted: November 25, 2001; Published: April 30, 2002

Citation of the article:

M. Petitjean, Solving the Geometric Docking Problem for Planar and Spatial Sets, *Internet Electron. J. Mol. Des.* 2002, 1, 185–192, <http://www.biochempress.com>.

Solving the Geometric Docking Problem for Planar and Spatial Sets[#]

Michel Petitjean*

ITODYS (CNRS, ESA 7086), 1 rue Guy de la Brosse, 75005 Paris, France

Received: October 16, 2001; Accepted: November 25, 2001; Published: April 30, 2002

Internet Electron. J. Mol. Des. 2002, 1 (4), 185–192

Abstract

Motivation. A docking algorithm working without charge calculations is needed for molecular modeling studies. Two sets of n points in the d -dimensional Euclidean space are considered. The optimal translation and/or rotation minimizing the variance of the sum of the n squared distances between the fixed and the moving set is computed. An analytical solution is provided for d -dimensional translations and for planar rotations. The use of the quaternion representation of spatial rotations leads to the solving of a quadratically constrained non-linear system. When both spatial translations and rotations are considered, the system is solved using a projected Lagrangian method requiring only 4-dimensional initial starting tuples.

Method. The projected Lagrangian method was used in the docking algorithm.

Results. The automatic positioning of the moving set is performed without any *a priori* information about the initial orientation.

Conclusions. Minimizing the variance of the squared distances is an original and simple geometric docking criterion, which avoids any charge calculation.

Availability. The FORTRAN source is available within framework of scientific collaborations. Contact: petitjean@itodys.jussieu.fr.

Keywords. Geometric docking; optimal rotation and translation; constrained optimization.

1 INTRODUCTION

The geometric docking problem comes from the molecular modeling field. Basically, two sets of punctual charges should face one to the other following some criterion. Since years, docking operations with energy and/or force feedback have been carried out by translating or rotating one molecule relative to another [11]. Charge and energy calculations being too time consuming, it was desirable to elaborate a purely geometric algorithm able to produce the desired optimal orientation. Let $\{\mathbf{x}0_i, i = 1, \dots, n\}$ and $\{\mathbf{x}1_i, i = 1, \dots, n\}$ be the two sets of n points in the d -dimensional Euclidean space. $\mathbf{X}0$ and $\mathbf{X}1$ are their respective associated (n, d) arrays. \mathbf{R} is a d -dimensional

[#] Dedicated on the occasion of the 70th birthday to Professor Alexandru T. Balaban.

* Correspondence author; phone: 33-(0)1-4427-4857; fax: 33-(0)1-4427-6814; E-mail: petitjean@itodys.jussieu.fr.

rotation and \mathbf{t} a d -dimensional translation. The transposed of a vector \mathbf{x} is denoted \mathbf{x}' , and the vector product of two vectors \mathbf{x} and \mathbf{y} is denoted $\mathbf{x} \wedge \mathbf{y}$. We consider the population of the n squares of the distances d_i :

$$d_i^2 = (\mathbf{R}(\mathbf{x}\mathbf{1}_i + \mathbf{t}) - \mathbf{x}\mathbf{0}_i)'(\mathbf{R}(\mathbf{x}\mathbf{1}_i + \mathbf{t}) - \mathbf{x}\mathbf{0}_i) \quad (1)$$

This population of n squared distances has a mean and a variance. The scope of this paper is to compute the optimal translation \mathbf{t} and rotation \mathbf{R} minimizing the variance. The pairwise correspondence between the members of the two sets $\{\mathbf{x}\mathbf{0}_i\}$ and $\{\mathbf{x}\mathbf{1}_i\}$ is assumed to be known, or is computed with an adequate algorithm [9].

Although the minimization of the mean is a well-known problem [3,6,8,10], it seems to be the first time that the variance minimization is considered. The major difficulty in these two constrained optimization problems comes from the condition that \mathbf{R} is a pure rotation rather than an orthogonal matrix. Minimizing the mean when \mathbf{R} is an orthogonal matrix is known as the orthogonal Procrustes problem, and has received a general solution using a singular value decomposition algorithm [5,7]. When \mathbf{R} is indeed a pure rotation, minimizing the mean has received solutions only for planar sets [8] and for spatial sets [3,6,10]. The solution for planar sets involves angles rather than 2D-rotation matrices, and the solution for spatial sets involves quaternions, either in their (4,4) matrix representation [3,6], or in their unit 4-vector representation [10]. The matricial expression of \mathbf{R} , subject to $\mathbf{R}'\mathbf{R} = \mathbf{I}$ and to $\det(\mathbf{R}) = +1$ has not been used, explaining why the pure rotation Procrustes problem has not received a general solution. For the same reasons, computing the rotation minimizing the variance has been done only for $d \leq 3$.

2 MATERIALS AND METHODS

2.1 The Optimal Translation

The rotation is constant. For clarity, \mathbf{R} is set to \mathbf{I} . The mean points are $\mathbf{g}\mathbf{0} = (\sum \mathbf{x}\mathbf{0}_i)/n$ and $\mathbf{g}\mathbf{1} = (\sum \mathbf{x}\mathbf{1}_i)/n$. Setting $\mathbf{e}_i = (\mathbf{x}\mathbf{1}_i - \mathbf{g}\mathbf{1}) - (\mathbf{x}\mathbf{0}_i - \mathbf{g}\mathbf{0})$ and $\boldsymbol{\tau} = \mathbf{t} + (\mathbf{g}\mathbf{1} - \mathbf{g}\mathbf{0})$, equation (1) becomes:

$$d_i^2 = (\mathbf{e}_i + \boldsymbol{\tau})'(\mathbf{e}_i + \boldsymbol{\tau}) \quad (2)$$

The mean of the population of the squared distances is:

$$\overline{d^2} = \boldsymbol{\tau}'\boldsymbol{\tau} + (\sum \mathbf{e}_i' \mathbf{e}_i) / n \quad (3)$$

The sum of the \mathbf{e}_i is null. The variance is:

$$V = (\sum d_i^4) / n - \overline{d^2}^2$$

$$V = \sum (\mathbf{e}_i' \mathbf{e}_i)^2 / n - (\sum \mathbf{e}_i' \mathbf{e}_i / n)^2 + 4 \sum (\mathbf{e}_i' \boldsymbol{\tau})^2 / n + 4 \sum (\mathbf{e}_i' \mathbf{e}_i) (\mathbf{e}_i' \boldsymbol{\tau}) / n$$

Let V_p be the initial variance prior translation, \mathbf{K} be the covariance matrix of the $\{\mathbf{e}_i\}$ population,

and \mathbf{I} be the third order term:

$$V_p = \Sigma(\mathbf{e}_i' \cdot \mathbf{e}_i)^2 / n - (\Sigma \mathbf{e}_i' \cdot \mathbf{e}_i / n)^2 \quad (4)$$

$$\mathbf{K} = \Sigma(\mathbf{e}_i \cdot \mathbf{e}_i') / n \quad (5)$$

$$\mathbf{l} = \Sigma(\mathbf{e}_i \cdot \mathbf{e}_i' \cdot \mathbf{e}_i) / n \quad (6)$$

Thus the variance is a quadratic expression of $\boldsymbol{\tau}$:

$$V = V_p + 4 \cdot \boldsymbol{\tau}' \cdot \mathbf{K} \cdot \boldsymbol{\tau} + 4 \cdot \mathbf{l}' \cdot \boldsymbol{\tau} \quad (7)$$

The optimal translation is get by solving the linear system:

$$2 \cdot \mathbf{K} \cdot \boldsymbol{\tau} + \mathbf{l} = 0 \quad (8)$$

Provided that the $\{\mathbf{e}_i\}$ population is fully d -dimensional, the optimal translation \mathbf{t} exists and is unique: $\mathbf{t} = -(\mathbf{g}\mathbf{1} - \mathbf{g}\mathbf{0} + \mathbf{K}^{-1} \cdot \mathbf{l} / 2)$. The minimized variance is:

$$V_m = V_p - \mathbf{l}' \cdot \mathbf{K}^{-1} \cdot \mathbf{l} \quad (9)$$

When the $\{\mathbf{e}_i\}$ population has a dimensionality $\delta < d$, the projection of $\boldsymbol{\tau}$ in the δ -subspace is computed first, and the projection of $\boldsymbol{\tau}$ in the orthogonal $(d - \delta)$ -subspace is free. This situation arises when $n \leq d$.

2.2 The Optimal Planar Rotation

The translation is constant. For clarity, \mathbf{t} is set to zero. The planar rotation is:

$$\mathbf{R} = \mathbf{I} \cdot \cos(r) + \mathbf{\Pi} \cdot \sin(r) \quad (10)$$

where \mathbf{I} is the identity matrix (*i.e.* the null rotation), and $\mathbf{\Pi}$ is the antisymmetric matrix associated to the +90 degrees rotation. Using (10), equation (1) is rewritten:

$$d_i^2 = \mathbf{x0}_i' \cdot \mathbf{x0}_i + \mathbf{x1}_i' \cdot \mathbf{x1}_i - 2 \cos(r) \cdot \mathbf{x0}_i' \cdot \mathbf{x1}_i - 2 \sin(r) \cdot \mathbf{x0}_i' \cdot \mathbf{\Pi} \cdot \mathbf{x1}_i \quad (11)$$

Let us define: $t_i = \mathbf{x0}_i' \cdot \mathbf{x0}_i + \mathbf{x1}_i' \cdot \mathbf{x1}_i$, $c_i = \mathbf{x0}_i' \cdot \mathbf{x1}_i$ and $s_i = \mathbf{x0}_i' \cdot \mathbf{\Pi} \cdot \mathbf{x1}_i$. Thus:

$$d_i^2 = t_i - 2 \cos(r) \cdot c_i - 2 \sin(r) \cdot s_i$$

The variance of the population $V = \text{Var}(\{d_i^2\})$ is:

$$\begin{aligned} V = & \text{Var}(\{t_i\}) + 4 \cos^2(r) \cdot \text{Var}(\{c_i\}) + 4 \sin^2(r) \cdot \text{Var}(\{s_i\}) \\ & - 4 \cos(r) \cdot \text{COV}(\{c_i\}, \{t_i\}) - 4 \sin(r) \cdot \text{COV}(\{s_i\}, \{t_i\}) \\ & + 8 \sin(r) \cos(r) \cdot \text{COV}(\{c_i\}, \{s_i\}) \end{aligned} \quad (12)$$

where COV is the covariance operator.

V_p being the variance prior rotation, equation (12) is rewritten:

$$\begin{aligned}
 V &= V_p - 4 \sin^2(r) \cdot \text{Var}(\{c_i\}) + 4 \sin^2(r) \cdot \text{Var}(\{s_i\}) \\
 &+ 4(1 - \cos(r)) \cdot \text{COV}(\{c_i\}, \{t_i\}) - 4 \sin(r) \cdot \text{COV}(\{s_i\}, \{t_i\}) \\
 &+ 8 \sin(r) \cos(r) \cdot \text{COV}(\{c_i\}, \{s_i\})
 \end{aligned}
 \tag{13}$$

Let us define: $CC = \text{Var}(\{c_i\})$, $SS = \text{Var}(\{s_i\})$, $CT = \text{COV}(\{c_i\}, \{t_i\})$, $ST = \text{COV}(\{s_i\}, \{t_i\})$, and $CS = \text{COV}(\{c_i\}, \{s_i\})$. The gradient is:

$$\frac{1}{4} \frac{\partial V}{\partial r} = -(CC - SS) \cdot \sin(2r) + (2 \cdot CS) \cdot \cos(2r) + CT \cdot \sin(r) - ST \cdot \cos(r)
 \tag{14}$$

Using $v = \text{tg}(r/2)$, the equation $\partial v / \partial r = 0$ leads to nullify the following quartic polynomial:

$$P(v) = (2CS + ST)v^4 + 2(2(CC - SS) + CT)v^3 - 6(2CS)v^2 + 2(-2(CC - SS) + CT)v + (2CS - ST)$$

V is a smooth periodic function of r , exhibiting therefore at least one minimum and one maximum over each period. Thus the quartic should have at least two real roots. It can be noticed that $P(0) + (P(1) + P(-1))/4$ is equal to the opposite of the v^4 coefficient. Therefore the quartic cannot have a constant sign and has indeed real roots. The roots are analytically computable [1].

2.3 The Optimal Spatial Rotation

The translation is constant. For clarity, \mathbf{t} is again set to zero. Using the unit 4–vector quaternionic representation of the spatial rotation, it is known (see appendix in ref. [10]) that equation (1) can be rewritten:

$$d_i^2 = d_{pi}^2 - 2\mathbf{q}'\mathbf{B}_i\mathbf{q}
 \tag{15}$$

where: $d_{pi}^2 = (\mathbf{x1}_i - \mathbf{x0}_i)'(\mathbf{x1}_i - \mathbf{x0}_i)$, and $\mathbf{B}_i = \begin{pmatrix} 0 & \mathbf{k}_i' \\ \mathbf{k}_i & \mathbf{A}_i \end{pmatrix}$, with $\mathbf{k}_i = \mathbf{x1}_i \wedge \mathbf{x0}_i$, and:

$$\mathbf{A}_i = (\mathbf{x1}_i \cdot \mathbf{x0}_i' + \mathbf{x0}_i \cdot \mathbf{x1}_i') - \text{Trace}(\mathbf{x1}_i \cdot \mathbf{x0}_i' + \mathbf{x0}_i \cdot \mathbf{x1}_i') \cdot \mathbf{I} = (\mathbf{x1}_i \cdot \mathbf{x0}_i' + \mathbf{x0}_i \cdot \mathbf{x1}_i') - 2(\mathbf{x0}_i' \cdot \mathbf{x1}_i) \cdot \mathbf{I}$$

and \mathbf{q} is the unknown unit quaternion.

Let us define: $\overline{d_p^2} = \Sigma d_{pi}^2 / n$, and $\overline{\mathbf{B}} = \Sigma \mathbf{B}_i / n$, and $\delta_i^2 = d_{pi}^2 - \overline{d_p^2}$, and $\mathbf{E}_i = 2(\mathbf{B}_i - \overline{\mathbf{B}})$. The variance operator being here insensitive to the terms that do not depend on the summation index, the variance $V = \text{Var}(\{d_i^2\})$ is rewritten: $V = \text{Var}(\{\delta_i^2 - \mathbf{q}'\mathbf{E}_i\mathbf{q}\})$. Let us define the symmetric matrices $\mathbf{M}_i = (\delta_i^2 \cdot \mathbf{I} - \mathbf{E}_i) / \sqrt{n}$, which are such that $\Sigma \mathbf{M}_i = 0$. Since $\mathbf{q}'\mathbf{q} = 1$, the variance to be minimized is now expressed as a sum of squared quadratic forms:

$$V = \Sigma (\mathbf{q}'\mathbf{M}_i\mathbf{q})^2
 \tag{16}$$

V is a 4th degree polynomial function of \mathbf{q} , which is to be minimized subject to $\mathbf{q}'\mathbf{q} = 1$. Furthermore, \mathbf{q} and $-\mathbf{q}$ are associated to the same rotation, and the first non–null component of the solution is to be set positive. When $n = 2$, $\mathbf{M}_2 = -\mathbf{M}_1$, and there is in fact only one quadratic form to

be minimized in modulus. When $n \geq 3$, no analytical solution appears. A numerical solver is required, such as the projected Lagrangian method [4]. Let L be the Lagrange multiplier associated to the quadratic constraint. The function to be minimized is:

$$F = \Sigma(\mathbf{q}'\mathbf{M}_i\mathbf{q})^2 - 2L(\mathbf{q}'\mathbf{q} - 1) \quad (17)$$

$$\mathbf{G} = 4(\Sigma(\mathbf{M}_i\mathbf{q}\mathbf{q}'\mathbf{M}_i\mathbf{q}) - L\mathbf{q}) \quad (18)$$

$$\mathbf{H} = 4(\Sigma((\mathbf{q}'\mathbf{M}_i\mathbf{q})\mathbf{M}_i + 2\mathbf{M}_i\mathbf{q}\mathbf{q}'\mathbf{M}_i) - L \cdot \mathbf{I}) \quad (19)$$

where \mathbf{G} and \mathbf{H} are respectively the gradient and the Hessian. The solutions of the system are noted \mathbf{q}^* , and are such that $\mathbf{G}^* = \mathbf{G}(\mathbf{q}^*) = 0$. Setting $\mathbf{q}^* \cdot \mathbf{G} = 0$ shows that the optimal value of the Lagrange multiplier is $L^* = F^*$, *i.e.* the optimized variance. Then, for each iteration, the value of the variance has been retained as an estimate of L . This value of L is such that

$$\frac{\partial(\mathbf{G}'\mathbf{G})}{\partial L} = 0$$

i.e. it minimizes $\|\mathbf{G}\|_2$ at each iteration. Each random starting \mathbf{q} value has been generated via normalization of a random 4–vector following the isotropic multinormal distribution and setting the first non–null component positive.

2.4 Solving the Full Docking Problem

Solving the full docking problem requires finding both the optimal translation and the optimal spatial rotation. According to equation (1), the translation is conventionally performed before the rotation. The analytical expression of the translation derived from equation (8) depends now on the unknown quaternion, and the 4–dimensional system derived from equations (17)–(19) depends on the unknown translation. The full system may be viewed as a quadratically constrained 7–dimensional non–linear system, ignoring the analytical solution of equation (8). But of course, a 4–dimensional system is desirable for obvious reasons, such as avoiding the use of numerous 7–dimensional starting points leading to uninteresting local minima. The full expression of the variance as a function of \mathbf{q} and \mathbf{t} is established hereafter. The analytical expression of d_i^2 and V given in equations (15) and (16) have to be updated, $\mathbf{x}\mathbf{1}_i$ being to be replaced by $\mathbf{x}\mathbf{1}_i + \mathbf{t}$. For convenience, the updated parameters are followed by $:=$. According to [2], a rotated vector \mathbf{z} is written:

$$\mathbf{R} \cdot \mathbf{z} = (\rho^2 - \mathbf{u}'\mathbf{u})\mathbf{z} + 2\mathbf{u} \cdot \mathbf{u}'\mathbf{z} + 2\rho \cdot \mathbf{u} \wedge \mathbf{z}$$

where ρ and \mathbf{u} are respectively the scalar part and the 3–vector part of \mathbf{q} , and are such that $\rho^2 + \mathbf{u}'\mathbf{u} = 1$. Then:

$$d_{pi}^2 := d_{pi}^2 + 2\mathbf{t}'(\mathbf{x}\mathbf{1}_i - \mathbf{x}\mathbf{0}_i) + \mathbf{t}'\mathbf{t}$$

$$\mathbf{k}_i := \mathbf{k}_i + \mathbf{t} \wedge \mathbf{x}\mathbf{0}_i$$

$$\mathbf{A}_i := \mathbf{A}_i + \mathbf{t} \cdot \mathbf{x0}_i' + \mathbf{x0}_i \cdot \mathbf{t}' - 2(\mathbf{x0}_i' \cdot \mathbf{t}) \cdot \mathbf{I}$$

Now, the centered values are defined: $\mathbf{y0}_i = \mathbf{x0}_i - \mathbf{g0}$ and $\mathbf{y1}_i = \mathbf{x1}_i - \mathbf{g1}$, $\mathbf{g0} = (\sum \mathbf{x0}_i)/n$ and $\mathbf{g1} = (\sum \mathbf{x1}_i)/n$ being the respective mean points. The matrices \mathbf{M}_i are now updated:

$$\mathbf{M}_i := \mathbf{M}_i + \frac{2}{\sqrt{n}} \mathbf{t}' (\mathbf{y1}_i - \mathbf{y0}_i) \cdot \mathbf{I} - \frac{2}{\sqrt{n}} \begin{pmatrix} 0 & (\mathbf{t} \wedge \mathbf{y0}_i)' \\ (\mathbf{t} \wedge \mathbf{y0}_i) & \mathbf{t} \cdot \mathbf{y0}_i' + \mathbf{y0}_i \cdot \mathbf{t}' - 2(\mathbf{y0}_i' \cdot \mathbf{t}) \cdot \mathbf{I} \end{pmatrix} \quad (20)$$

It follows from (20) than $\mathbf{q}'\mathbf{M}_i\mathbf{q}$ is updated by addition of a linear function of \mathbf{t} , which can be expressed as the dot product of \mathbf{t} by a suitable vector \mathbf{b}_i depending on \mathbf{q} :

$$\mathbf{q}'\mathbf{M}_i\mathbf{q} := \mathbf{q}'\mathbf{M}_i\mathbf{q} + \mathbf{b}_i' \cdot \mathbf{t} \quad (21)$$

$$\mathbf{b}_i = \frac{2}{\sqrt{n}} (\mathbf{y1}_i - \mathbf{y0}_i) - \frac{4}{\sqrt{n}} (\rho(\mathbf{y0}_i \wedge \mathbf{u}) + \mathbf{u} \cdot \mathbf{u}' \mathbf{y0}_i - \mathbf{y0}_i \cdot \mathbf{u}' \mathbf{u}) \quad (22)$$

The centered values $\mathbf{y0}_i$ and $\mathbf{y1}_i$ have a null sum, thus the vectors \mathbf{b}_i and the updated matrices \mathbf{M}_i have also a null sum. The final expression of the variance is:

$$V = \Sigma(\mathbf{q}'\mathbf{M}_i\mathbf{q} + \mathbf{b}_i' \cdot \mathbf{t})^2 \quad (23)$$

The variance has to be minimized in \mathbf{q} and \mathbf{t} , subject to $\mathbf{q}'\mathbf{q} = 1$. As in equation (17), the objective function to minimize is: $F = V - 2L(\mathbf{q}'\mathbf{q} - 1)$. Writing the Newton step, the unknown increment $\begin{pmatrix} \tilde{\mathbf{q}} \\ \tilde{\mathbf{t}} \end{pmatrix}$ satisfies:

$$\begin{pmatrix} \mathbf{G}_q \\ \mathbf{G}_t \end{pmatrix} + \begin{pmatrix} \mathbf{H}_{qq} & \mathbf{H}_{qt} \\ \mathbf{H}_{tq} & \mathbf{H}_{tt} \end{pmatrix} \cdot \begin{pmatrix} \tilde{\mathbf{q}} \\ \tilde{\mathbf{t}} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (24)$$

where \mathbf{G}_q and \mathbf{G}_t are the gradients respectively associated to \mathbf{q} and \mathbf{t} , \mathbf{H}_{qq} and \mathbf{H}_{tt} are the Hessians respectively associated to \mathbf{q} and \mathbf{t} , \mathbf{H}_{qt} is the rectangular (4,3) matrix of the cross derivatives and \mathbf{H}_{tq} is the transposed of \mathbf{H}_{qt} . Only \mathbf{G}_q and \mathbf{H}_{qq} are depending on L . Clearly, applying the optimal translation at each iteration of the 4–dimensional minimizing procedure described in section 3 would be equivalent to solving the system (24) with a block diagonal 7–dimensional Hessian, for which \mathbf{H}_{qt} and \mathbf{H}_{tq} are replaced by blocks of zeroes. But unfortunately, \mathbf{H}_{qt} is not null, even for $\mathbf{t} = 0$ and $\mathbf{R} = \mathbf{I}$:

$$\mathbf{G}_i = 2\Sigma \mathbf{b}_i \mathbf{q}' \mathbf{M}_i \mathbf{q} + 2(\Sigma \mathbf{b}_i \mathbf{b}_i') \mathbf{t} \quad (25)$$

$$\mathbf{G}_q = 4\Sigma(\mathbf{M}_i \mathbf{q} \mathbf{q}' \mathbf{M}_i \mathbf{q}) - 4L\mathbf{q} + 2\Sigma(\mathbf{q}' \mathbf{M}_i \mathbf{q}) (\Gamma_i \cdot \mathbf{t}) + 4\Sigma \mathbf{M}_i \mathbf{q} \mathbf{b}_i' \cdot \mathbf{t} + 2\Sigma(\mathbf{b}_i' \cdot \mathbf{t}) (\Gamma_i \cdot \mathbf{t}) \quad (26)$$

$$\mathbf{H}_{qt} = 2\Sigma(\mathbf{q}' \mathbf{M}_i \mathbf{q}) \Gamma_i + 4\Sigma(\mathbf{M}_i \mathbf{q} \mathbf{b}_i' + 2\Sigma(\mathbf{b}_i' \cdot \mathbf{t}) \mathbf{g}_i + 2\Sigma \Gamma_i \mathbf{t} \mathbf{b}_i') \quad (27)$$

where Γ_i is the (4,3) array of the gradient of \mathbf{b}_i in respect to \mathbf{q} . Even if the block diagonal system derived from (24) would keep some local quadratic convergence property, it should lead to a minimum depending on the initial translation. Thus, it is better to extract \mathbf{t} from the linear system \mathbf{G}_t

= 0, and report its value in equation (23). The variance V is now a function of \mathbf{q} only, and its minimum is get from solving a 4–dimensional problem subject to $\mathbf{q}'\mathbf{q} = 1$. The function to be minimized is:

$$F = \Sigma(\mathbf{q}'\mathbf{M}_i\mathbf{q})^2 - 2L(\mathbf{q}'\mathbf{q} - 1) - (\Sigma\mathbf{b}_i\mathbf{q}'\mathbf{M}_i\mathbf{q})(\Sigma\mathbf{b}_i\mathbf{b}_i')^{-1}(\Sigma\mathbf{b}_i\mathbf{q}'\mathbf{M}_i\mathbf{q}) \quad (28)$$

Now, it is pointed out that the first and second derivatives are all computed with $O(n)$ loops at each iteration. On an other hand, performing indeed the translation and the rotation associated to an iteration is also done in an $O(n)$ loop, and leads to simpler expressions of the derivatives. In this situation, $\mathbf{t} = 0$, $\rho = 1$, $\mathbf{u} = 0$, and $\mathbf{q} = (1,0,0,0)'$. Then, $\Sigma\mathbf{b}_i\mathbf{q}'\mathbf{M}_i\mathbf{q} = 0$. The function and the gradient are computable respectively by setting $\mathbf{q} = (1,0,0,0)'$ in equations (17) and (18). Computing the Hessian is performed by setting $\mathbf{q} = (1,0,0,0)'$ in equation (19) and adding a supplementary term $\Delta\mathbf{H}$. Setting $\mathbf{e}_i = \mathbf{y}\mathbf{1}_i - \mathbf{y}\mathbf{0}_i$ as in section 2, then $\mathbf{b}_i = 2\mathbf{e}_i / \sqrt{n}$, and $\mathbf{b}_i\mathbf{b}_i' = 4\mathbf{K}$, \mathbf{K} being the covariance matrix defined in equation (5). Then, the first line of $\mathbf{\Gamma}_i$ is null, and the (3,3) remaining block of $\mathbf{\Gamma}_i$ is $(-4/\sqrt{n}) \cdot \mathbf{Y}\mathbf{0}_i$, $\mathbf{Y}\mathbf{0}_i$ being the antisymmetric matrix build from $\mathbf{y}\mathbf{0}_i$, such that $\mathbf{Y}\mathbf{0}_i \cdot \mathbf{z} = \mathbf{y}\mathbf{0}_i \wedge \mathbf{z}$ for any 3–vector \mathbf{z} . The first diagonal element of \mathbf{M}_i is noted f_i . It follows that $\mathbf{q}'\mathbf{M}_i\mathbf{q} = f_i$ and then $\Sigma f_i \mathbf{e}_i = 0$. Now, the 3–vector \mathbf{m}_i is defined from the first column of \mathbf{M}_i : $\mathbf{m}_i' = [\mathbf{M}_i(2,1), \mathbf{M}_i(3,1), \mathbf{M}_i(4,1)]$, and the expression of the supplementary term $\Delta\mathbf{H}$ being to add to the Hessian is:

$$\Delta\mathbf{H} = -\frac{8}{n} \begin{pmatrix} 0 & 0 \\ 0 & \gamma_i' \mathbf{K}^{-1} \gamma_i \end{pmatrix} \quad (29)$$

$$\gamma_i = \Sigma \mathbf{e}_i \mathbf{m}_i' - \Sigma f_i \cdot \mathbf{Y}\mathbf{0}_i \quad (30)$$

3 RESULTS AND DISCUSSION

Solving the geometric docking problem needs to find the absolute minima for the variance V of the population of the n squared distances between n couples of points in R^3 . V is a real function of 7 parameters: 3 for the translation \mathbf{t} and 4 for the quaternion \mathbf{q} associated to the rotation. The constraint $\mathbf{q}'\mathbf{q}=1$ means that there are in fact 6 degrees of freedom. The variance V measures the dispersion around the mean value of the squared distances, and V is invariant when some constant is added to the squared distances. Thus, V can be nullified when $n = 7$. When $n < 7$, an infinite number of (\mathbf{q}, \mathbf{t}) values lead to $V = 0$. When $n > 7$, V has a finite number of minima. The non–linear system is solved by an iterative procedure, and leads therefore to a local minimum depending of an initial value. A part of the system being analytically solvable, the final solving procedure needs only initial \mathbf{q} values.

It has been observed that the efficiency of the projected Lagrangian procedure described in section 2.4 is highly dependant on its implementation. It is little sensitive to n , provided that $n > 7$. For $n = 7$, the efficiency was lower. Practical n values were up to some hundreds. The local

quadratic convergence was indeed obtained, and the final absolute minimum was each time checked using thorough Monte–Carlo experiments. The docking problem in the plane is solved analytically here, but docking in dimension higher than 3 is currently unsolved. It basically comes from that, compared to quaternions, derivatives and constraints associated to rotation matrices are more difficult to handle, formally and numerically.

4 CONCLUSIONS

Although minimizing the mean of the squared distances is a well known technique for spatial superpositions, (see [9] and references cited), minimizing the variance of the squared distances seems to be an original docking criterion. The present method offers various advantages in the context of molecular docking. The graphs of the molecules are not used, and time–consuming charge calculations are not needed. Providing a pertinent initial relative orientation of the molecules is not required here, and this is an original feature of our method.

5 REFERENCES

- [1] E. J. Barbeau, *Polynomials*, Springer–Verlag, New–York, 1989, section 1.4.
- [2] R. Deheuvels, *Formes quadratiques et groupes classiques*, Presses Universitaires de France, Paris, 1981, p 375.
- [3] R. Diamond, A Note on the Rotational Superposition Problem, *Acta Cryst. A* **1988**, *44*, 211–216.
- [4] P. E. Gill, W. Murray, M. H. Wright, *Practical Optimization*, Academic Press, London, 1981, sections 6.5 and 6.6.
- [5] G. H. Golub, C. F. Van Loan, *Matrix Computations*, John Hopkins University Press, Baltimore, 1985, section 12.4.
- [6] B. K. P. Horn, Closed–Form Solution of Absolute Orientation using Unit Quaternions, *J. Opt. Soc. Am. A* **1987**, *4*, 629–642.
- [7] J. R. Hurley, R. B. Cattell, The Procrustes Program: Producing Direct Rotation to Test a Hypothesized Factor Structure, *Behav. Sci.* **1962**, *7*, 258–262.
- [8] M. Petitjean, About Second Kind Continuous Chirality Measures. 1. Planar Sets, *J. Math. Chem.* **1997**, *22*, 185–201.
- [9] M. Petitjean, Interactive Maximal Common 3D Substructure Searching with the Combined SDM/RMS Algorithm, *Comp. Chem.* **1998**, *22*, 463–465.
- [10] M. Petitjean, On the Root Mean Square Quantitative Chirality and Quantitative Symmetry Measures, *J. Math. Phys.* **1999**, *40*, 4587–4595.
- [11] TRIPOS Inc.: Biopolymer Manual, Version 6.6, 1699 South Hanley Road St. Louis, MO 63144: TRIPOS Inc. **1999**, section 3.1.