STARDUST SCHOOL

Roma Tor Vergata 8–12 September 2014

Four Lectures on Impact Monitoring

Andrea Milani

Department of Mathematics

University of Pisa, Italy

The challenge of the STARDUST project is to communicate with mathematicians, aerospace engineers and astronomers, as if they were a single community, with the same language, customs and interests.

Plan Lecture 1: Linearized Impact Probability

1. **The Problem of Orbit Determination**

2. **Nonlinear Least Squares**

3. **Gaussian random variables**

4. **Probabilistic Interpretation of Orbit Determination**

5. **Target planes: linear and semilinear predictions**

6. **Linearized Impact Monitoring**

1.1 Orbits [1.1]

The three elements of an orbit determination problem are:

1. orbits, 2. observations and 3. error model.

Orbits are solutions of an equation of motion:

$$\frac{d\mathbf{y}}{dt} = \mathbf{f}(\mathbf{y}, t, \boldsymbol{\mu})$$

which is an ordinary differential equation; $\mathbf{y} \in \mathbb{R}^p$ is the **state vector**, $\boldsymbol{\mu} \in \mathbb{R}^{p'}$ are the **dynamical parameters**, such as the masses of the planets, $t \in \mathbb{R}$ the time.

The **initial conditions** are the value of the state vector at an epoch t_0: $\mathbf{y}(t_0) = \mathbf{y_0} \in \mathbb{R}^p$. All the orbits together form the general solution

$$\mathbf{y} = \mathbf{y}(t, \mathbf{y_0}, \boldsymbol{\mu}) \ .$$

1.2 Observations [1.1]

For the second element we introduce an **observation function**

$$r = R(\mathbf{y}, t, \mathbf{v})$$

depending on the current state, directly upon time, and upon a number of **kinematical parameters** $\mathbf{v} \in \mathbb{R}^{p''}$. The function R is differentiable. The composition of the general solution with the observation function is the **prediction function**

$$r(t) = R(\mathbf{y}(t), t, \mathbf{v})$$

which is used to predict the outcome of a specific observation at some time t_i, with $i = 1, \ldots, m$. However, the observation result r_i is generically not equal to the prediction, the difference being the **residual**

$$\xi_i = r_i - R(\mathbf{y}(t_i), t_i, \mathbf{v}) \quad , \quad i = 1, \ldots, m \ .$$

The observation function can depend also upon the index i, e.g., when using a 2-dimensional observation function either (right ascension, declination) or (range, range-rate). All the residuals can be assembled forming a vector in $\mathbb{R}^m$

$$\boldsymbol{\xi} = (\xi_i)_{i=1,\ldots,m}$$

which is in principle a function of all the $p + p' + p''$ variables $(\mathbf{y_0}, \boldsymbol{\mu}, \mathbf{v})$.

1.3 Errors and Target Function [1.1-1.2]

The random element is introduced by the assumption that **every observation contains an error**. Even if we know all the true values $(\mathbf{y_0}^*, \boldsymbol{\mu}^*, \mathbf{v}^*)$ of the parameters and our explicit computations are perfectly accurate, nevertheless the residuals

$$\xi_i^* = r_i - R(\mathbf{y}(\mathbf{y_0}^*, t_i, \boldsymbol{\mu}^*), t_i, \mathbf{v}^*, i) = \varepsilon_i$$

would not be zero but random variables. The joint distribution of $\boldsymbol{\varepsilon} = (\varepsilon_i)_{i=1,\dots,m}$ needs to be modeled, either in the form of a probability density function or as a set of inequalities, describing the observation errors we rate as acceptable.

The basic tool of the classical theory of orbit determination [Gauss 1809] is the definition of a **target function** $Q(\boldsymbol{\xi})$ depending on the vector of residuals $\boldsymbol{\xi}$. The target function needs suitable conditions of regularity and convexity. We shall focus on the simplest case, with Q proportional to the **sum of squares** of all the residuals

$$Q(\boldsymbol{\xi}) = \frac{1}{m} \boldsymbol{\xi}^T \boldsymbol{\xi} = \frac{1}{m} \sum_{i=1}^{m} \xi_i^2 \; .$$

Since each residual is a function of all the parameters,

$$\xi_i = \xi_i(\mathbf{y_0}, \boldsymbol{\mu}, \mathbf{v}) \; ,$$

the target function is also a function of $(\mathbf{y_0}, \boldsymbol{\mu}, \mathbf{v})$.

1.4 The Least Squares Principle [1.2]

The next step is to select the **parameters to be fit to the data**: let $\mathbf{x} \in \mathbb{R}^N$ be a sub-vector of $(\mathbf{y_0}, \boldsymbol{\mu}, \boldsymbol{\nu}) \in \mathbb{R}^{p+p'+p''}$, that is $\mathbf{x} = (x_i), i = 1, N$ with each x_i either a component of the initial conditions, or a dynamical parameter, or a kinematic parameter. Then we consider the target function

$$Q(\mathbf{x}) = Q(\boldsymbol{\xi}(\mathbf{x}))$$

as a function of $\mathbf{x}$ only, leaving the **consider parameters** $\mathbf{k} \in \mathbb{R}^{p+p'+p''-N}$ (all the parameters not included in $\mathbf{x}$) fixed at the assumed value.

The **minimum principle** selects as **nominal solution** the point $\mathbf{x}^* \in \mathbb{R}^N$ where the target function $Q(\mathbf{x})$ has its minimum value Q^* (at least a local minimum). The principle of **least squares** is the minimum principle with as target function the sum of squares $Q(\boldsymbol{\xi}) = \boldsymbol{\xi}^T \boldsymbol{\xi}/m$ (or some other quadratic form).

1.5 The Optimization Interpretation [1.3]

The minimum principle should not be understood as if the "real" solution needs to be the point of minimum $\mathbf{x}^*$. Two interpretations can be used.

According to the **optimization interpretation**, $\mathbf{x}^*$ is the optimum point but values of the target function immediately above the minimum are also acceptable. The set of acceptable solutions can be described as the **confidence region**

$$Z(\sigma) = \left\{ \mathbf{x} \in \mathbb{R}^N \;\middle|\; \sum_{i=1}^{N} \xi_i^2 \leq m Q^* + \sigma^2 \right\}$$

depending upon the **confidence parameter** $\sigma > 0$. The solutions $\mathbf{x}$ in $Z(\sigma)$ correspond to observation errors larger that those for $\mathbf{x}^*$, but still compatible with the quality of the observation procedure. The choice of the value of σ bounding the acceptable errors is not easy.

The alternative **probabilistic interpretation** describes the observation errors ε_i as random variables with an assumed probability density.

2.1 Nonlinear Least squares [5.2]

The target function of the nonlinear **least squares** problem

$$Q(\mathbf{x}) = \frac{1}{m} \, \boldsymbol{\xi}^T(\mathbf{x}) \, \boldsymbol{\xi}(\mathbf{x})$$

is a differentiable function of the **fit parameters** $\mathbf{x}$, not just a quadratic function. The partial derivatives of the residuals with respect to the fit parameters are

$$B = \frac{\partial \boldsymbol{\xi}}{\partial \mathbf{x}}(\mathbf{x}) \quad , \quad H = \frac{\partial^2 \boldsymbol{\xi}}{\partial \mathbf{x}^2}(\mathbf{x})$$

where the **design matrix** B is an $m \times N$ matrix, with $m \geq N$, H is a 3-index array of shape $m \times N \times N$. The partials of the residuals are computed by the chain rule:

$$\frac{\partial \xi_i}{\partial x_k} = -\frac{\partial R}{\partial \mathbf{y}} \frac{\partial \mathbf{y}(t_i)}{\partial x_k} - \frac{\partial R}{\partial x_k}$$

by using the first term if x_k belongs to $(\mathbf{y_0}, \boldsymbol{\mu})$ (initial condition/dynamical), the second if x_k belongs to $\boldsymbol{\nu}$ (kinematic). The formula for H is complicated.

To find the minimum, we look for stationary points of $Q(\mathbf{x})$:

$$\frac{\partial Q}{\partial \mathbf{x}} = \frac{2}{m} \, \boldsymbol{\xi}^T B = \mathbf{0} \ .$$

2.2 Iterative methods: Newton [5.2]

The standard **Newton method** involves the computation of the second derivatives of the target function:

$$\frac{\partial^2 Q}{\partial \mathbf{x}^2} = \frac{2}{m}\left(B^T B + \xi^T H\right) = \frac{2}{m} C_{new} \tag{1}$$

where C_{new} is a $N \times N$ matrix. Given the residuals $\xi(\mathbf{x}_k)$ at iteration k, the (non-zero) gradient is expanded around $\mathbf{x}_k$ and equated to zero:

$$\mathbf{0} = \frac{\partial Q}{\partial \mathbf{x}}(\mathbf{x}) = \frac{\partial Q}{\partial \mathbf{x}}(\mathbf{x}_k) + \frac{\partial^2 Q}{\partial \mathbf{x}^2}(\mathbf{x}_k)\,(\mathbf{x} - \mathbf{x}_k) + \dots$$

where the dots stand for terms of higher order in $(\mathbf{x} - \mathbf{x}_k)$. Thus

$$C_{new}\,(\mathbf{x}^* - \mathbf{x}_k) = -B^T\,\xi + \dots$$

Neglecting the higher order terms, if the matrix C_{new}, as computed at the point $\mathbf{x}_k$, is invertible then the Newton iteration $k+1$ provides a correction $\mathbf{x}_k \longrightarrow \mathbf{x}_{k+1}$ with

$$\mathbf{x}_{k+1} = \mathbf{x}_k + C_{new}^{-1}D \quad, \quad D = -B^T\,\xi.$$

The point $\mathbf{x}_{k+1}$ should be a better approximation to $\mathbf{x}^*$ than $\mathbf{x}_k$. In practice, the Newton method may converge or not, depending upon the **first guess** $\mathbf{x}_0$ selected to start the iterations; if it converges, the limit is a stationary point of $Q(\mathbf{x})$.

2.3 Iterative methods: Differential Corrections [5.2]

The most used method is a variant of the Newton method, known in this context as **differential corrections**, with each iteration making the correction

$$\mathbf{x}_{k+1} = \mathbf{x}_k - (B^T\, B)^{-1}\, B^T \boldsymbol{\xi}$$

where the **normal matrix** $C = B^T\, B$, computed at $\mathbf{x}_k$, replaces C_{new}. One **iteration of differential correction** is the solution of a linearized least squares problem, with **normal equation**

$$C\left(\mathbf{x}_{k+1} - \mathbf{x}_k\right) = D$$

where the right hand side $D = -B^T\,\boldsymbol{\xi}$ is the same as in the Newton method. Thus, if the iterations converge, the limit point is a stationary point of $Q(\mathbf{x})$: the only stationary point which cannot be reached are the ones which are not local minima. The use of higher derivatives does not remove the need for a good first guess $\mathbf{x}_0$.

This linearized problem can be obtained by the truncation of the target function

$$Q(\mathbf{x}) \simeq Q(\mathbf{x}_k) + \frac{2}{m}\,\boldsymbol{\xi}^T\, B\left(\mathbf{x} - \mathbf{x}_k\right) + \frac{1}{m}\left(\mathbf{x} - \mathbf{x}_k\right)^T C\left(\mathbf{x} - \mathbf{x}_k\right),$$

which is not the full Taylor expansion to order 2, since C_{new} is replaced by C. We neglect in the normal equation, on top of the terms of order ≥ 2 in $(\mathbf{x}^* - \mathbf{x}_k)$, also the term $\boldsymbol{\xi}^T\, H\left(\mathbf{x}^* - \mathbf{x}_k\right)$, which is of the first order in $(\mathbf{x}^* - \mathbf{x}_k)$ but contains also $\boldsymbol{\xi}$, thus is smaller than $C\left(\mathbf{x}^* - \mathbf{x}_k\right)$.

2.4 Propagation of Normal and Covariance matrices [5.3]

For the normal matrix at the initial time t_0:

$$C_0 = \frac{\partial \boldsymbol{\xi}}{\partial \mathbf{y_0}}^T \frac{\partial \boldsymbol{\xi}}{\partial \mathbf{y_0}}$$

the propagation to time t is obtained by assuming the fit variables are $\mathbf{y}(t)$, then by applying to the **state transition matrix** the **chain rule**:

$$C_t = \frac{\partial \boldsymbol{\xi}}{\partial \mathbf{y}}^T \frac{\partial \boldsymbol{\xi}}{\partial \mathbf{y}} = \left(\frac{\partial \boldsymbol{\xi}}{\partial \mathbf{y_0}} \frac{\partial \mathbf{y_0}}{\partial \mathbf{y}} \right)^T \left(\frac{\partial \boldsymbol{\xi}}{\partial \mathbf{y_0}} \frac{\partial \mathbf{y_0}}{\partial \mathbf{y}} \right) = \left[\left(\frac{\partial \mathbf{y}}{\partial \mathbf{y_0}} \right)^{-1} \right]^T C_0 \left(\frac{\partial \mathbf{y}}{\partial \mathbf{y_0}} \right)^{-1}$$

The covariance matrices are the inverse of the normal matrices, thus

$$\Gamma_0 = C_0^{-1} \quad , \quad \Gamma_t = C_t^{-1} = \frac{\partial \mathbf{y}}{\partial \mathbf{y_0}} \Gamma_0 \frac{\partial \mathbf{y}}{\partial \mathbf{y_0}}^T \quad ,$$

giving the **covariance propagation** formula, the same as in the probabilistic interpretation. To propagate the normal and covariance matrix it is not necessary to solve again the least square problem, but only to solve the variational equation, which provides the state transition matrix $\partial \mathbf{y}/\partial \mathbf{y_0}$.

3.1 Gaussian random variables [3.1-3.2]

A **continuous random variable** X is defined by a **probability density function (PDF)** $p_X(x) \geq 0$ for $x \in \mathbb{R}$ with the property $\int_{-\infty}^{+\infty} p_X(x)\, dx = 1$. The probability for X to be in the open interval (a, b) is $\mathcal{P}_X(a < X < b) = \int_a^b p_X(x)\, dx$.

There are continuous random variables playing an important role in the least squares principle, those with probability density function of the type

$$p_X(x) = N(\mu, \sigma^2)(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

where $\mu = E(X) = \int_{-\infty}^{+\infty} x\, p_X(x)\, dx$ is the expected value, $\sigma = \mathrm{STD}(X)$ is the square root of the variance $\sigma^2 = Var(X) = \int_{-\infty}^{+\infty} [x - E(X)]^2\, p_X(x)\, dx$. Such random variables are called **Gaussian** or **normally distributed**.

3.2 Rotational invariance and Gaussian [3.2]

For two jointly distributed random variables X, Y with PDF $p_{X,Y}(x,y)$ the **marginal PDF**

$$p_X(x) = \int_{-\infty}^{+\infty} p_{X,Y}(x,y)\, dy \quad , \quad p_Y(y) = \int_{-\infty}^{+\infty} p_{X,Y}(x,y)\, dx,$$

is the PDF of one of the two, valid for all possible values of the other variable. X, Y are **independent random variables** if $p_{X,Y}(x,y) = p_X(x)\, p_Y(y)$.

A geometric characterization:

if the random variables X, Y are **independent**, with equal **marginal densities** $p_X(x) = p_Y(x) = f(x)$ and the PDF $p_{X,Y}(x,y)$ is **invariant by rotation**, i.e. there is a function $g : \mathbb{R} \to \mathbb{R}$ such that $p_{X,Y}(x,y) = g(x^2 + y^2)$, then **they are Gaussian with zero mean**: $p_X(x) = N(0, \sigma^2)(x)$.

3.3 Multidimensional Gaussians [3.2]

Given n jointly distributed random variables $X_1, X_2, \ldots, X_n$, we say that they are **Gaussian**, or **normally distributed**, if their joint PDF is of the form

$$p_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = \frac{\sqrt{\det C}}{(2\pi)^{n/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{m})^T C (\mathbf{x} - \mathbf{m})\right]$$

where $\mathbf{m} = (m_1, \ldots, m_n)^T$ is the vector of the means and the normal matrix C is symmetric and positive definite.

The notation $p_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y}) = N(\mathbf{m}, \Gamma)$ is used for the PDF above, where $\Gamma = C^{-1}$ is the **covariance matrix**. Of the coefficients of $\Gamma = (\gamma_{ik})$, the diagonal ones are $\gamma_{ii} = Var(x_i) = STD(x_i)^2$ and the **correlations** are contained in the off-diagonal elements of Γ:

$$\gamma_{ik} = Cov(x_i, x_k) = STD(x_i) STD(x_k) Corr(x_i, x_k)$$

If the **normal matrix** C is diagonal so is Γ, and the X_j are all independent: for Gaussian variables, being independent and uncorrelated is equivalent.

3.4 Marginal Gaussians [3.2]

For multidimensional Gaussian, we need to generalize the result on **marginal PDF**: let us consider the two vector random variables

$$\mathbf{X} = (X_1, \ldots, X_n) \quad , \quad \mathbf{Y} = (Y_1, \ldots, Y_m)$$

jointly distributed, Gaussian with probability density

$$p_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y}) = N\left((\mathbf{m_x}; \mathbf{m_y}), \Gamma\right),$$

where $\mathbf{x} = (x_1, \ldots, x_n)$, $\mathbf{y} = (y_1, \ldots, y_m)$, $(\mathbf{m_x}; \mathbf{m_y})$ is the stacking of the two vectors, and the covariance matrix can be decomposed as

$$\Gamma = \begin{bmatrix} \Gamma_{\mathbf{x}} & \Gamma_{\mathbf{xy}} \\ \Gamma_{\mathbf{yx}} & \Gamma_{\mathbf{y}} \end{bmatrix} \quad , \quad \Gamma_{\mathbf{yx}} = \Gamma_{\mathbf{xy}}^T,$$

where $\Gamma_{\mathbf{x}}$ is $n \times n$, $\Gamma_{\mathbf{y}}$ is $m \times m$ and $\Gamma_{\mathbf{xy}}$ is $n \times m$. If $\Pi_{\mathbf{x}}$ is the matrix of the projection on the $\mathbf{x}$ subspace, $\Gamma_{\mathbf{x}} = \Pi_{\mathbf{x}} \Gamma \Pi_{\mathbf{x}}^T$, similarly for $\Gamma_{\mathbf{y}}$.

The **marginal probability densities**

$$p_{\mathbf{X}} = N(\mathbf{m_x}, \Gamma_{\mathbf{x}}) \quad , \quad p_{\mathbf{Y}} = N(\mathbf{m_y}, \Gamma_{\mathbf{y}}),$$

are such that the **marginal covariance matrix** is the restriction, to the corresponding linear subspace, of the covariance matrix. The **marginal normal matrices** are $C^{\mathbf{x}} = \Gamma_{\mathbf{x}}^{-1}, C^{\mathbf{y}} = \Gamma_{\mathbf{y}}^{-1}$.

3.5 Conditional Gaussians [3.2-3.3]

Given X,Y with PDF $\mathsf{p}_{X,Y}(x,y)$ the **conditional probability density** are

$$p_{X|Y}(x;y) = \frac{\mathsf{p}_{X,Y}(x,y)}{\mathsf{p}_Y(y)} \quad , \quad p_{Y|X}(y;x) = \frac{\mathsf{p}_{X,Y}(x,y)}{\mathsf{p}_X(x)} ,$$

for $\mathsf{p}_Y(y) > 0$ and $\mathsf{p}_X(x) > 0$, respectively.

For a multivariate Gaussian PDF $\mathsf{p}_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) = N(\mathbf{m},\Gamma)$

$$\mathbf{p}_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}:\mathbf{y}) = N(\mathbf{m_x} + \Gamma_{\mathbf{xy}}\Gamma_{\mathbf{y}}^{-1}(\mathbf{y}-\mathbf{m_y}), \Gamma_{\mathbf{x}} - \Gamma_{\mathbf{xy}}\Gamma_{\mathbf{y}}^{-1}\Gamma_{\mathbf{yx}}) , \qquad (2)$$

which can be described as follows: the **conditional normal matrix** $C_{\mathbf{X}}$ is the restriction, to the corresponding linear subspace, of the normal matrix $C = \Gamma^{-1}$. The **conditional covariance matrix** is $\Gamma^{\mathbf{x}} = (C_{\mathbf{x}})^{-1}$. Similarly for $\mathbf{p}_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}:\mathbf{x})$.

The equation $\mathbf{x} = \mathbf{m_x} + \Gamma_{\mathbf{xy}}\Gamma_{\mathbf{y}}^{-1}(\mathbf{y}-\mathbf{m_y})$ defines a linear space, the **regression subspace**, containing for each $\mathbf{y}$ the expected values of $\mathbf{p}_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}:\mathbf{y})$.

3.6 Example in 2 dimensions [5.4]

Given **two Gaussian random variables g and h**, the joint PDF has level manifolds which are ellipsoids (ellipses for $N = 2$). The **level manifolds of the marginal PDF are projections**, the **conditional PDF are intersections**:
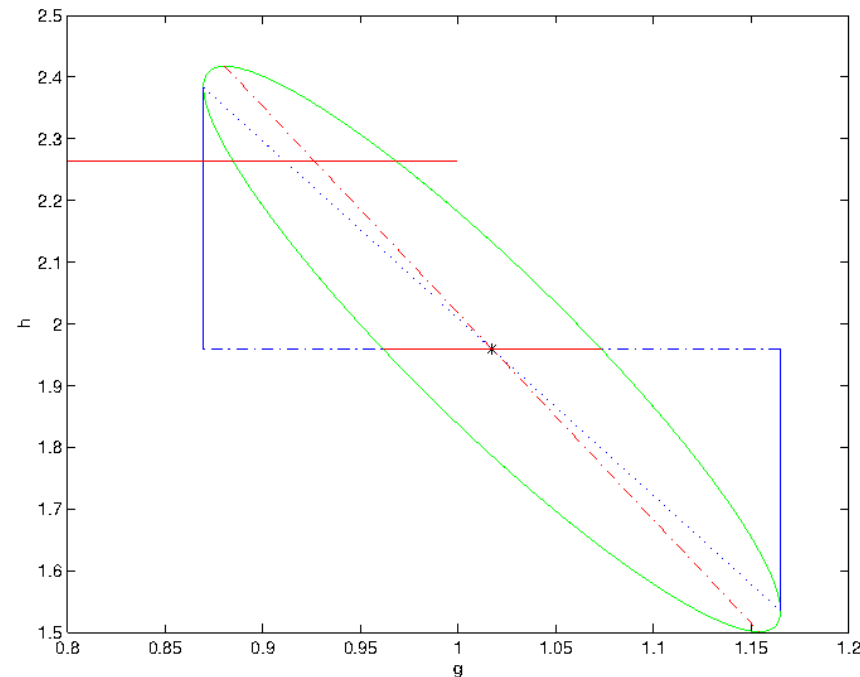
The **regression line of g given h** (dash dot) contains the centers of the $\mathbf{h} = const$ sections, in this $N = 2$ case the midpoint of the horizontal intersection segments. The **regression line of h given g** (dotted) contains the midpoints of the vertical intersections segments, including the points of tangency with vertical lines.

3.7 Propagation of Gaussians [3.3]

Let $\mathbf{X}$ be a normal variable, with PDF $p_{\mathbf{X}}(\mathbf{x}) = N(\mathbf{m}, \Gamma)$, $\mathbf{m} \in \mathbb{R}^n$ and Γ a symmetric positive definite $n \times n$ matrix. Let $\mathbf{y} = \mathbf{f}(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$, with A an invertible $n \times n$ matrix, be an affine transformation in $\mathbb{R}^n$. Then $\mathbf{Y} = \mathbf{F}(\mathbf{X})$ also has a normal distribution, with PDF

$$p_{\mathbf{Y}}(\mathbf{y}) = N\left(A\mathbf{m} + \mathbf{b}, A\,\Gamma\,A^T\right),$$

that is, with expected value $\mathbf{f}(\mathbf{m})$ and covariance matrix $A\,\Gamma\,A^T$. This is called the **covariance propagation** rule, following from the change of variable formula for multiple integrals.

A **generalization to transformations** $\mathbf{y} = \mathbf{f}(\mathbf{x}) = B\mathbf{x} + \mathbf{b}$ of the Gaussian variable $\mathbf{X}$ with PDF $N(\mathbf{m}, \Gamma)$, where B is an $m \times n$ matrix $(m < n)$ with maximal rank m, can be obtained as follows:

$$p_{\mathbf{Y}}(\mathbf{y}) = N\left(B\mathbf{m} + \mathbf{b}, B\Gamma B^T\right).$$

4.1 The simplest Probabilistic Error Model [5.7]

The **probabilistic interpretation** uses as source of randomness the residuals $\boldsymbol{\xi}$. The simplest assumption is that, after the best possible value has been found for the fit parameters $\mathbf{x}$, each residual ξ_i is a continuous random variable Ξ_i with **zero mean and unit variance** (in some appropriate unit), independent from the index i. It is also assumed that the error of each observation is a random variable **independent** from the ones of the other observations.

Under the additional hypothesis that the **joint PDF is continuous and rotation invariant**, i.e., a function depending only upon the sum of squares, that is upon the target function $Q(\mathbf{x})$, the only possible PDF is the **Gaussian** one $p_{\Xi_i}(\xi_i) = N(0,1)(\xi_i)$. Then the residuals random vector $\boldsymbol{\xi}$ has probability density

$$p_{\boldsymbol{\Xi}}(\boldsymbol{\xi}) = N(\mathbf{0}, I)(\boldsymbol{\xi})$$

with I the $m \times m$ identity matrix.

Under these conditions the vector solution for the fit parameters $\mathbf{x}$ can be seen as a set of jointly distributed random variables $\mathbf{X}$: the goal is **to compute the PDF** $p_{\mathbf{X}}(\mathbf{x})$, **given the probability density** $p_{\boldsymbol{\Xi}}(\boldsymbol{\xi})$.

4.2 The Manifold of Possible Residuals [5.7]

The **residuals are a function of the fit parameters** (and of the observations):

$$G : \mathbb{R}^N \longrightarrow \mathbb{R}^m \quad , \quad \boldsymbol{\xi} = G(\mathbf{x})$$

obtained by subtracting from the observations the prediction function. Let $\mathbf{x}^*$ be the **nominal solution** and $\boldsymbol{\xi}^* = G(\mathbf{x}^*)$ the corresponding residuals. G is a differentiable function, thus we can linearize at the nominal solution

$$\boldsymbol{\xi} - \boldsymbol{\xi}^* = B(\mathbf{x}^*) (\mathbf{x} - \mathbf{x}^*) + \dots$$

where $B(\mathbf{x}^*)$ is the **design matrix**, computed at convergence of the differential corrections, and the dots stand for terms of order higher than 1 in $|\mathbf{x} - \mathbf{x}^*|$.

The **image of the fit parameters space $V = G(\mathbb{R}^N)$ is an N–dimensional submanifold of the residuals space $\mathbb{R}^m$**. This manifold can have singularities, but the point $\boldsymbol{\xi}^*$ cannot be singular, because the matrix $B(\mathbf{x}^*)$ has rank N, otherwise differential corrections would fail and the nominal solution $\mathbf{x}^*$ could not be reached. Thus we can assume that the manifold V is smooth, at least near $\boldsymbol{\xi}^*$.

4.3 The linear space of Residuals [5.7]

We need to compute the **conditional PDF** of Ξ on V, as a step to compute the PDF of $\mathbf{X}$ on $\mathbb{R}^N$. If we neglect the higher order terms we get a linearized equation

$$\Delta\boldsymbol{\xi} = B(\mathbf{x}^*)\,\Delta\mathbf{x} \quad, \quad \Delta\boldsymbol{\xi} = \boldsymbol{\xi} - \boldsymbol{\xi}^* \quad, \quad \Delta\mathbf{x} = \mathbf{x} - \mathbf{x}^*,$$

the tangent map between $\mathbb{R}^N$ and the N**-plane** $TV(\boldsymbol{\xi}^*)$ **tangent to** V **at** $\boldsymbol{\xi}^*$.

To use this linearization is as considering **the linear least squares problem with quadratic target function** $Q(\mathbf{x}) = Q(\mathbf{x}^*) + \frac{1}{m}\,(\mathbf{x} - \mathbf{x}^*)^T\,C\,(\mathbf{x} - \mathbf{x}^*)$ neglecting all higher order terms (by using C instead of C_{new} we neglect also the $\boldsymbol{\xi}^T H$ term).

In this approximation, we can compute the PDF of Ξ on $TV(\boldsymbol{\xi}^*)$. This is actually a **conditional PDF** for a Gaussian PDF $N(\mathbf{0}, I)$ under the assumption $\Xi \in TV(\boldsymbol{\xi}^*)$. If we model $TV(\boldsymbol{\xi}^*)$ as follows:

$$TV(\boldsymbol{\xi}^*) = \{\boldsymbol{\xi} \in \mathbb{R}^m : \boldsymbol{\xi} = B\,(\mathbf{x} - \mathbf{x}^*) + \boldsymbol{\xi}^*, \mathbf{x} \in \mathbb{R}^N\} \; ;$$

we can use a rotation matrix R such that

$$R(\boldsymbol{\xi} - \boldsymbol{\xi}^*) = \begin{bmatrix} \boldsymbol{\xi}' \\ \boldsymbol{\xi}'' \end{bmatrix} \Longrightarrow R^T \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\xi}'' \end{bmatrix} + \boldsymbol{\xi}^* \in TV(\boldsymbol{\xi}^*) \,,$$

that is, $\boldsymbol{\xi}'' \in \mathbb{R}^N$ **parameterizes** $TV(\boldsymbol{\xi}^*)$.

4.4 Propagation of PDF to possible residuals [3.3-5.7]

The PDF $N(\mathbf{0}, I)$ is **rotation invariant**, thus the PDF of Ξ'' can be computed as **conditional PDF on the Ξ'' subspace**. Thus Ξ'' has a Gaussian PDF with as normal matrix the restriction of the normal matrix of Ξ:

$$p_{\Xi''} = N\left(\mathbf{0}, I^{-1}\right) = N(\mathbf{0}, I)$$

with I the $N \times N$ identity matrix. Geometrically, the intersection of $(m-1)$-spheres with N-planes can only be $(N-1)$-spheres, and these are the level surfaces of the probability density of Ξ''.

In these coordinates the linearized map $B(\mathbf{x}^*)$ has a **simpler structure**, since the ξ' component of the image is $\mathbf{0}$:

$$R B(\mathbf{x}^*) = \begin{bmatrix} \mathbf{0} \\ A \end{bmatrix}$$

with $A = A(\mathbf{x}^*)$ an invertible $N \times N$ matrix. Then the normal matrix $C(\mathbf{x}^*) = B^T(\mathbf{x}^*) B(\mathbf{x}^*)$ is

$$\mathbf{C} = B^T B = B^T R^T R B = \mathbf{A^T A} .$$

4.5 Propagation of PDF to the solution [3.3-5.7]

The inverse transformation from $TV(\boldsymbol{\xi}^*)$ to $\mathbb{R}^N$ is given by the matrix A^{-1}: by the **Gaussian propagation** formula for the invertible case, the PDF of $\mathbf{X}$

$$p_{\mathbf{X}}(\mathbf{x}) = N\left(\mathbf{x}^*, A^{-1} I [A^{-1}]^T\right) = N(\mathbf{x}^*, \Gamma)$$

is Gaussian with covariance matrix

$$\Gamma = A^{-1} [A^{-1}]^T = [A^T A]^{-1} = [B^T B]^{-1} = C^{-1}$$

Some of the fundamental results from the book **Gauss (1809)**:

- The solution of a linear least squares problem has a Gaussian PDF, with mean equal to the nominal $\mathbf{x}^*$ and covariance matrix Γ.

- $\Gamma = C^{-1}$ is the matrix solving the normal equation, thus connecting the differential corrections with both probabilistic and optimization interpretation.

- The computations required by the optimization and by the probabilistic interpretation are the same: the result is defined by $\mathbf{x}^*$, $C(\mathbf{x}^*)$, $\Gamma(\mathbf{x}^*)$, Q^*.

The Gaussian PDF of the solution contains the residuals only through the penalty function $\Delta Q(\mathbf{x}) = Q(\mathbf{x}) - Q(\mathbf{x}^*)$, in the quadratic approximation for $Q(\mathbf{x})$

$$p_{\mathbf{X}}(\mathbf{x}) = \frac{\sqrt{\det C}}{(2\pi)^{N/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T C (\mathbf{x} - \mathbf{x}^*)\right].$$

4.6 Weighting and debiasing [5.3]

A simple generalization of the least squares problem is the **weighted least squares problem**, with a general positive definite quadratic form as target function:

$$Q(\boldsymbol{\xi}) = \frac{1}{m} \left(\boldsymbol{\xi} - \mathbf{b}\right)^T W \left(\boldsymbol{\xi} - \mathbf{b}\right) = \frac{1}{m} \sum_{i=1}^{m} \sum_{k=1}^{m} w_{ik} \left(\xi_i - b_i\right) \left(\xi_k - b_k\right)$$

where $W = (w_{ik})$ is the **weight matrix**, a symmetric matrix with positive eigenvalues, and the **bias vector** $\mathbf{b} = (b_i)$ corrects for systematic errors. This corresponds to the second order approximation to a generic function with a local minimum.

The bias $\mathbf{b}$ can be interpreted as the expected value and W as the normal matrix of the Gaussian distribution of $\boldsymbol{\xi}$, which is $N(\mathbf{b}, W^{-1})$.

The normal equation changes: the matrix C and the right hand side D contain W

$$C = B^T W B \quad , \quad D = -B^T W \left(\boldsymbol{\xi} - \mathbf{b}\right) .$$

4.7 Weighting and debiasing [5.3-5.7]

Non-uniform weights and biases can be formally introduced by using a **normalization of the residuals**. The **Cholewsky algorithm** is a procedure to find an upper triangular matrix P such that $P^T P = W$. Then we can obtain the **normalized residuals** $\boldsymbol{\xi}$ from the true residuals $\boldsymbol{\xi}'$ by

$$\boldsymbol{\xi} = P\left(\boldsymbol{\xi}' - \mathbf{b}\right)$$

These normalized residuals all have distribution $N(0,1)$ and are independent. The matrix of partial derivatives B of $\boldsymbol{\xi}$ is obtained from the matrix of partial derivatives B' of $\boldsymbol{\xi}'$

$$
\begin{aligned}
B &= \frac{\partial \boldsymbol{\xi}}{\partial \mathbf{x}} = \frac{\partial \boldsymbol{\xi}}{\partial \boldsymbol{\xi}'}\frac{\partial \boldsymbol{\xi}'}{\partial \mathbf{x}} = P\,B'\,, \\
C &= (B')^T\,W\,B' = (B')^T\,P^T\,P\,B' = B^T\,B\,, \\
D &= -(B')^T\,W\,\boldsymbol{\xi}' = -(B')^T\,P^T\,P\,\boldsymbol{\xi}' = -B^T\,\boldsymbol{\xi}
\end{aligned}
$$

and the weight matrix again disappears from the normal equation. Thus we may use the formulas in which the weight matrix W does not appear but still **assume that the observations have been weighted and debiased**.

5.1 Target Planes: MTP [12.1]

The geometry of the encounters with a planet can be described in terms of a **target plane** in 3-D space containing the **center of mass (CoM)** of the planet, orthogonal to the direction of the relative velocity of the approaching small body. Then an **impact** is an orbit containing a target plane point inside the planet cross section.

There are two ways to define such a target plane. The simplest is the **modified target plane (MTP)**: it is obtained by considering the time $\bar{t}$ at which the small body orbit has a relative minimum of the distance from the planet CoM.

Let $\mathbf{d}$ and $\mathbf{v}$ be the planetocentric position and velocity vectors at the time $\bar{t}$: the distance being minimum, $\mathbf{d} \cdot \mathbf{v} = 0$. The MTP is the plane, containing $\mathbf{0}$ (the CoM) and normal to $\mathbf{v}$. On this plane the point $\mathbf{d}$ is the **MTP close approach trace**.

The **MTP impact cross section** of the planet is a disk centered at $\mathbf{0}$ and with the radius R of the planet; if the minimum distance is $d = |\mathbf{d}| < R$ there is an impact.

5.2 Target Planes: TP [12.1]

The other definition, called either just **target plane (TP)** or **b-plane**, uses the same vectors $\mathbf{d}$ and $\mathbf{v}$ at the closest approach time $\bar{t}$ to compute a planetocentric 2-body approximation of the orbit.

If, as it is generally the case, such 2-body orbit is **hyperbolic**, then the TP is the plane containing $\mathbf{0}$ and **orthogonal to the incoming asymptote of the hyperbola**, corresponding to the limit vector $\mathbf{u}$ for $t \to -\infty$ of the planetocentric velocity. The size $u = |\mathbf{u}|$ is the **velocity "at infinity"** V_∞ as used in Astrodynamics.

The point $\mathbf{b}$, representing the **TP close approach trace**, is the intersection of the asymptote with the TP; its length $b = |\mathbf{b}|$ is the **impact parameter**, larger than the minimum distance d by a factor

$$\frac{b}{d} = \sqrt{\frac{v^2 d}{v^2 d - 2\,GM}}$$

where GM is the gravitational active mass.

On the TP the **impact cross section** is a disk of radius $B = R\sqrt{1 + 2\,GM/Ru^2}$ larger than the radius R by a factor accounting for gravitational focusing.

5.3 Target Planes: Coordinates [12.1]

A complete description of the close approaching orbit is obtained by assigning **two coordinates** ξ, ζ **on the TP**, two angles θ, ϕ defining **the orientation of the TP**, the size of the **velocity "at infinity"** $u = |\mathbf{u}|$ and the **time** $\bar{t}$.

The two planes are different, because the velocity $\mathbf{v}$ at the close approach is **rotated by an angle** $\gamma/2$ around the axis of the planetocentric angular momentum. The **angle** γ **measures the total deflection** from the incoming to the outgoing asymptote and can be computed by

$$\sin(\gamma/2) = \frac{GM}{v^2\, d - GM} \ .$$

The transformation of coordinates rotating and rescaling the MTP into the TP is **not canonical**; in fact, it is **impossible to use the Hamiltonian formalism including coordinates on the TP** (Tommei 2006).

From an abstract point of view, **it does not matter how we select a representative vector** for a given close approach, provided it is a **smooth function** of the orbit initial conditions. However, some coordinate systems are more equal than others, because **the propagation of the uncertainty is easy in a linear approximation**.

6.1 Search for Future Target Planes [12.1]

For a given asteroid with initial condition $\mathbf{x} \in \mathbb{R}^6$ at epoch t_0 there is a unique orbit, which can be accurately propagated for some time span, e.g., 100 years. For **each close approach to the Earth**, occurring within this time span, there is at least one **point $\mathbf{y} \in \mathbb{R}^2$ which is the trace of this orbit on the target plane**.

To avoid geometric complications, we consider as **close approach** only an encounter with a distance from the planet CoM **not exceeding some value D**; practical values for D range between 0.05 and 0.2 Astronomical Units (AU), thus the target planes are replaced by disks with a finite radius[*].

Let the orbit determination solve only for the initial conditions $\mathbf{x} \in \mathbb{R}^6$ at some epoch t_0, and the differential corrections converge to the nominal solution $\mathbf{x}^*$, with normal and covariance matrices C, Γ. As the nominal solution $\mathbf{x}^*$ is surrounded by a **6-dimensional confidence region of acceptable solutions**, the trace point $\mathbf{y}^* = \mathbf{g}(\mathbf{x}^*)$ determined by the propagated nominal orbit on the target plane of some encounter is surrounded by a **2-dimensional confidence region**.

---

[*]It is possible for a close approach to have multiple local minima of the distance to the CoM, thus multiple target plane trace points. Reducing D can often eliminate such complications.

6.2 Linear Prediction on Target Planes [12.1]

To compute a linear approximation, we use the **differential of the map $\mathbf{g}(\mathbf{x})$ providing the TP trace**. The trace point is reached at the time $t_c(\mathbf{x})$ of the **target plane crossing** for each orbit with initial conditions $\mathbf{x}$ in a neighborhood of $\mathbf{x}^*$.

In Cartesian geocentric coordinates ξ, η, ζ such that $\eta = 0$ is the target plane, the **equation $\eta(t, \mathbf{x}) = 0$, with $d\eta/dt > 0$, implicitly defines** the crossing time $t_c(\mathbf{x})$ as a differentiable function thus $\xi(t_c(\mathbf{x}), \mathbf{x})$ and $\zeta(t_c(\mathbf{x}), \mathbf{x})$ are differentiable too.

Using the differential $D\mathbf{g}(\mathbf{x}^*) = \partial(\xi, \zeta)(\mathbf{x}^*)/\partial\mathbf{x}$, a 2×6 matrix, we can compute the covariance matrix of the $\mathbf{y}$ prediction by the linear **covariance propagation** formula

$$\Gamma_{\mathbf{y}} = D\mathbf{g}\,\Gamma_{\mathbf{x}}\,(D\mathbf{g})^T \quad , \quad C_{\mathbf{y}} = \Gamma_{\mathbf{y}}^{-1}$$

defining the **confidence ellipse** on the target plane

$$(\mathbf{y} - \mathbf{y}^*)^T C_{\mathbf{y}}\,(\mathbf{y} - \mathbf{y}^*) \leq \sigma^2$$

with the same **confidence parameter** σ used for the confidence ellipsoids. If this quadratic approximation is adequate, the **possibility of an impact** can be studied by looking for intersections of the confidence ellipses with the impact cross section.

6.3 Linear Computation of Impact Probability [12.1]

By the Gaussian probabilistic formalism, from the normal probability density $N(\mathbf{x}^*, \Gamma)$ we can **propagate a PDF $\mathbf{P_Y}(\mathbf{y})$ on the target plane**. In the linear approximation with the differential $D\mathbf{g}(\mathbf{x}^*)$, $\mathbf{Y}$ is Gaussian with density $\mathbf{P_Y}(\mathbf{y}) = N(\mathbf{y}^*, \Gamma_{\mathbf{y}})$. Then it is possible to estimate the **Impact Probability** by computing a **probability integral on the impact cross section**, which is $E = \{|\mathbf{y}| < B\}$ if we are using the TP

$$\mathcal{P}(\mathbf{y} \in E) = \int\int_E \mathbf{P_Y}(\mathbf{y}) \, d\mathbf{y}$$

The formalism above is well known for the applications to **the navigation of interplanetary spacecraft**, a case in which the assumptions of small confidence regions, thus the applicability of linearization are well founded.

To **estimate the probability of impact of asteroids is much more difficult**, due to strong non-linearity.

Plan Lecture 2: Line Of Variations method for impact monitoring.

Engineers and mathematicians are different, but they share the lack of interest for deep philosophical questions, for which they cannot find a source of absolute truth. As an example, the question: "down to which probability you wish to be informed of a possible impact?" is not easy to handle for both of them.

7. **Why linear impact monitoring does not work? 1997 XF11**

8. **Nonlinear propagation of uncertainty**

9. **Definition of Line Of Variation LOV**

10. **Practical computation of the LOV**

11. **Non-uniqueness of the LOV**

7.1 The 1997 XF11 case: "It is zero, folks"

On 11/03/1998 the Minor Planet Center (official IAU repository for astrometric/orbital information on asteroids) announced (e-mail circular and press release) that asteroid 1997 XF11 had the possibility of impacting the Earth in 2028. In few hours, the JPL dynamics group announced that the Impact Probability (IP) in 2028 was effectively zero (> 10 standard deviations).
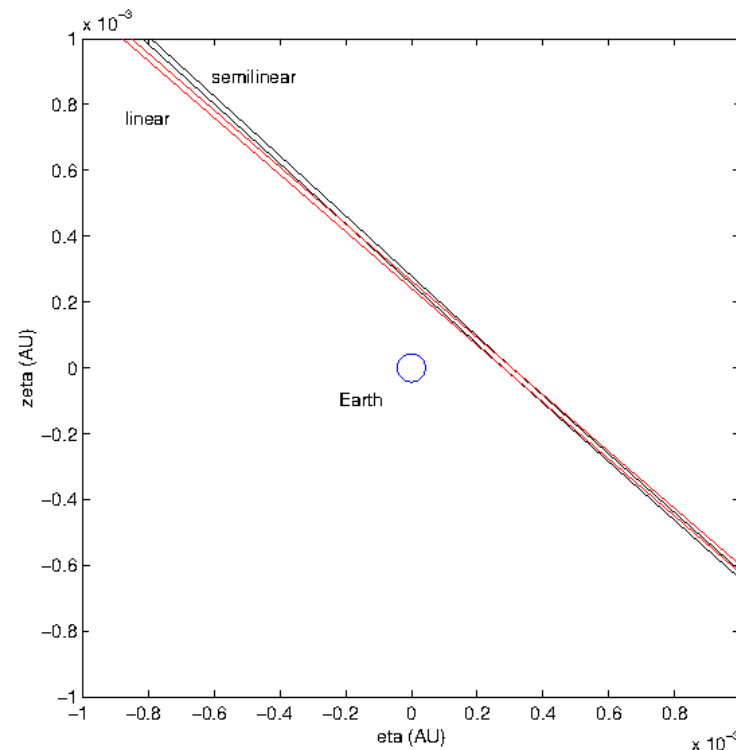
Few hours later, a precovery of 1997 XF11 was found on a 1990 photographic plate. Thus in 2028 there was going to be no collision: minimum distance ~ 1 Million km.

7.2 The 1997 XF11 case: the whole truth

On 12/03/1998 G.B. Valsecchi and myself started working on the problem: our "post mortem" (submitted 26/06/1998) of the 1997 XF11 PR catastrophe concluded that both sides of the dispute were wrong. MPC, because the impact in 2028 was never possible. JPL, because they had used an IP computed by a linear approximation.
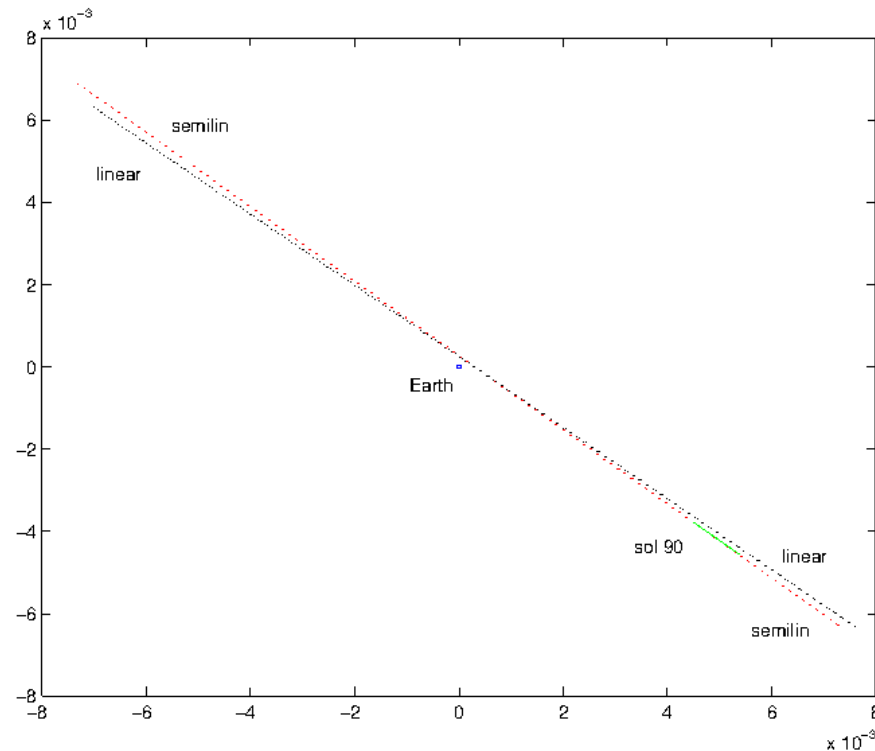
Linear confidence ellipse (3 STD) on the 2028 MTP and semilinear confidence boundary. Near Earth, the difference is small, farther out the difference is large. The solution with 1990 data is outside the ellipse, inside the semilinear curve.

7.3 The 1997 XF11 case: how we could fail

An enlargement near the tip of the confidence boundaries shows that there the linear prediction and the semilinear one are incompatible. Of course the semilinear is better.
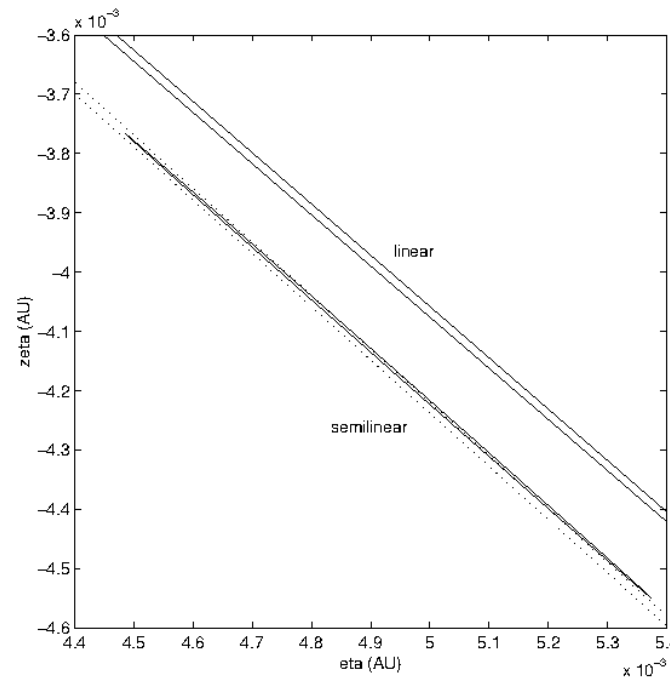
Indeed, as pointed out by MPC, with the 1997-98 data only, for a successive close approach in 2040 an impact was possible, although this cannot be found by a linearized theory. Indeed, with the LOV method we were soon able to compute an Impact Probability (IP) of $1/22,000$ for 26 October 2040.

7.4 The paradox of Impact Probability and Non-linearity

Can the linear theory be used to compute IP? The linear theory is always applicable near the nominal solution: if the entire $3 - STD$ boundary ellipse is a few $10,000$ km, that is a few Earth radii, the linear theory is a good approximation. To the contrary, if the confidence boundary is $\sim 1\,M$ km long, the higher order terms in the target function $Q(\mathbf{x})$ (of degree ≥ 3 in $\mathbf{x} - \mathbf{x}^*$) are relevant and a linear prediction can be wrong both ways, i.e., giving a possible impact when it is impossible as well as missing an impact possibility.

Then the choice of the method depends upon the IP we are interested in. If the IP is of the order of 1, e.g., > 0.5, the Earth impact cross section contains a good fraction of the $3 - STD$ confidence ellipse, with longest axis no more than a few Earth radii: thus the linear computation is a good approximation. If the IP $< 1/10,000$ then, in most cases, the longest axis of the $3 - STD$ confidence ellipse has to be many $1,000$ Earth radii, and the linear formula does not apply at all.

For the planning of Gravity Assist Maneuvers, which need to be successful with probability ~ 0.99, the navigation engineers are perfectly right in using a linear theory. For Impact Monitoring in case of a possible Extinction Level Event (like 1997 XF11: 1.6 km diameter), a probability of $1/10,000,000$ is by no means irrelevant, then "It's zero folks" from a linear computation does not imply it is safe.

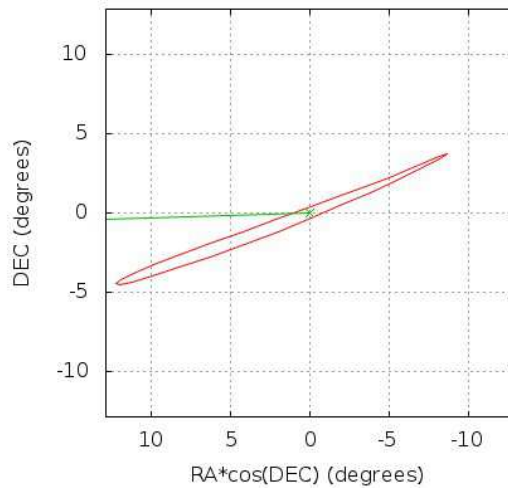# 8.1 Nonlinear propagation of the uncertainty: an example

Uncertainty Region for Asteroid 2011CG42
(relative to the nominal observation)
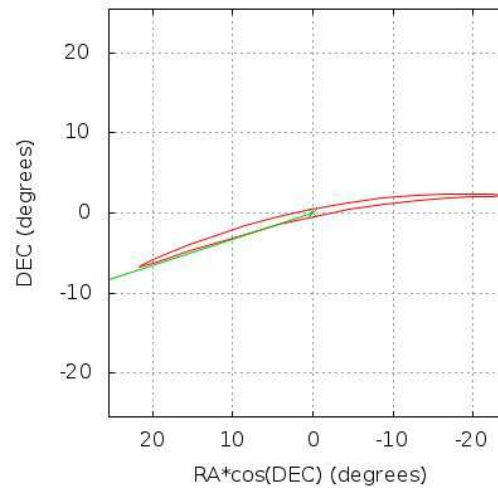Date and time 2011/04/01 15:58:00 UTC, Sigma level 3.0
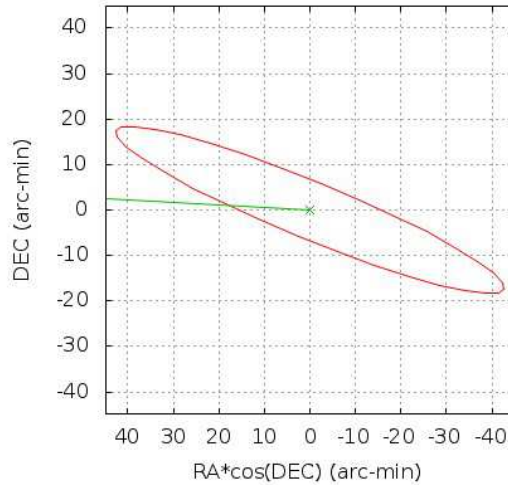
Uncertainty Region for Asteroid 2011CG42
(relative to the nominal observation)
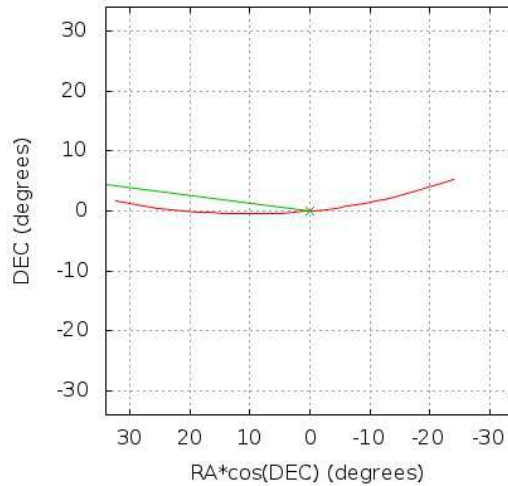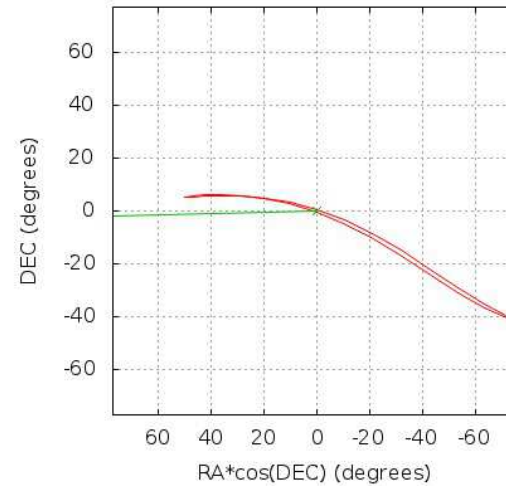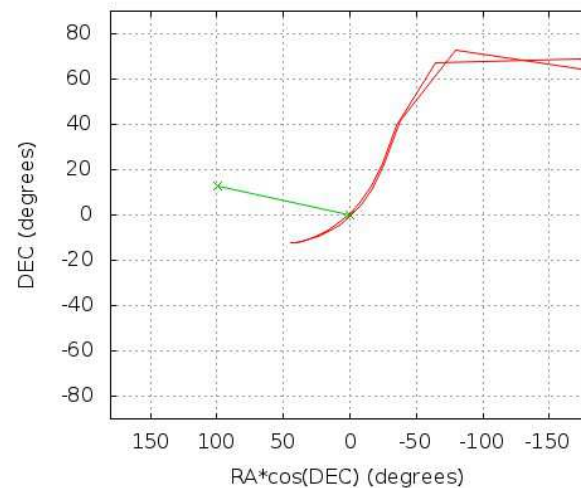Date and time 2011/09/01 15:58:00 UTC, Sigma level 3.0
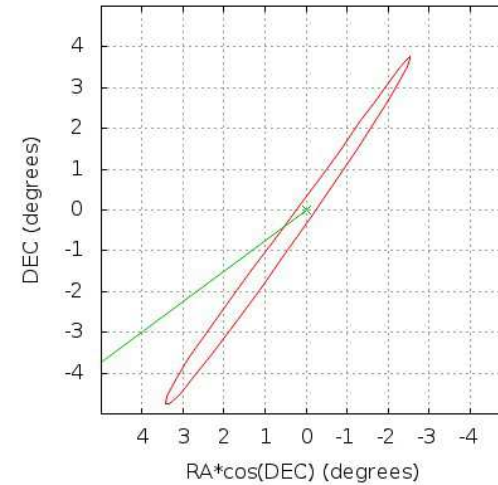
Uncertainty Region for Asteroid 2011CG42
(relative to the nominal observation)
Date and time 2011/12/01 16:02:00 UTC, Sigma level 3.0

Uncertainty Region for Asteroid 2011CG42
(relative to the nominal observation)
Date and time 2012/09/01 16:03:00 UTC, Sigma level 3.0

8.2 Nonlinear propagation of the uncertainty: an example

Uncertainty Region for Asteroid 2011CG42
(relative to the nominal observation)
Date and time 2013/03/01 16:04:00 UTC, Sigma level 3.0

Uncertainty Region for Asteroid 2011CG42
(relative to the nominal observation)
Date and time 2013/09/01 16:04:00 UTC, Sigma level 3.0

Uncertainty Region for Asteroid 2011CG42
(relative to the nominal observation)
Date and time 2014/03/01 16:07:00 UTC, Sigma level 3.0

Uncertainty Region for Asteroid 2011CG42
(relative to the nominal observation)
Date and time 2014/09/01 16:08:00 UTC, Sigma level 3.0
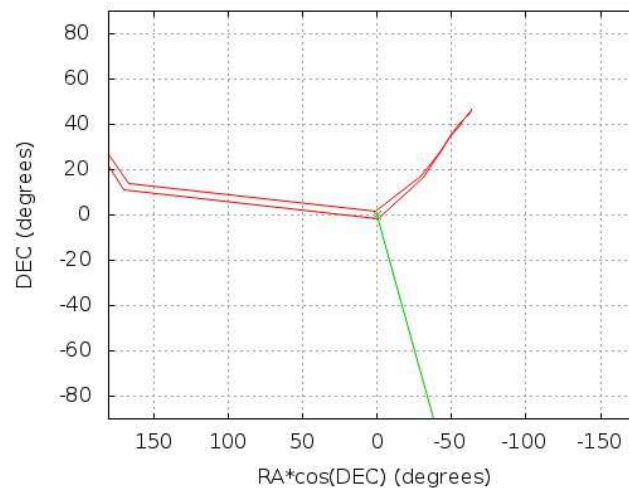
9.1 Complete Rank Deficiency [6.1]

If, at some step $\mathbf{x}_j$ of the differential corrections, the normal matrix C is not invertible, then the correction solving

$$B^T(\mathbf{x}_j)\,B(\mathbf{x}_j)\,(\mathbf{x}_{j+1}-\mathbf{x}_j) = -B^T(\mathbf{x}_j)\,\boldsymbol{\xi}(\mathbf{x}_j)$$

cannot be computed by means of the covariance matrix Γ. Solutions of the normal equation anyway exist (but are not unique). The pseudo-inverse C^*, is the matrix associated to the null map on the kernel of C times the inverse of C restricted to the subspace orthogonal to the kernel; C^* provides the solution of minimum norm

$$\mathbf{x}_{j+1} = \mathbf{x}_j - C^*(\mathbf{x}_j)\,B^T(\mathbf{x}_j)\,\boldsymbol{\xi}(\mathbf{x}_j)\ .$$

The pseudo-inverse C^* can be used as generalized covariance matrix for some purposes. However, corrections based on the pseudo-inverse are unlikely to converge towards a minimum of $Q(\mathbf{x})$.

Let us suppose that there is a rank deficiency of order $d = 1$, not just in one point, but over a large portion of the $\mathbf{x}$ space*. What would happen if we were to use an iteration with C^* replacing the non-existing Γ? Can this iteration converge?

*This can occur as a consequence of including in $\mathbf{x}$ two functionally dependent parameters.

9.2 Differential corrections with singular normal matrix

If the iteration step uses the pseudo-inverse

$$\mathbf{x}_{j+1} = \mathbf{x}_j + C^* D \quad ; \quad D = -B^T \boldsymbol{\xi}$$

then at each step there are no changes in $\mathbf{x}$ along the Weak direction, defined by the eigenvector $\mathbf{v}_1$ of the eigenvalue 0 of C.

If in some point of the $\mathbf{x}$ space $D(\mathbf{x})$ is parallel to $\mathbf{v}_1(\mathbf{x})$, there is no change. Thus if the sequence is convergent $\lim_{j \to \infty} \mathbf{x}_j = \mathbf{x}^\circ$, then $D(\mathbf{x}^\circ)$ is parallel to $\mathbf{v}_1(\mathbf{x}^\circ)$.

Definition: The points $\mathbf{x}$ such that $D(\mathbf{x})$ is parallel to $\mathbf{v}_1(\mathbf{x})$ belong to the Line Of Variation (LOV). The equations imposing parallelism are $N - 1$ if $\mathbf{x} \in \mathbb{R}^N$, thus by the implicit function theorem the LOV is generically a 1-dimensional manifold, possibly with singular points, that is a line.

Numerically, zero eigenvalues may disappear, e.g., a normal matrix C with eigenvalues $0 = \lambda_1 < \lambda_2 \leq \ldots \leq \lambda_N$ if computations are done in exact arithmetic can appear with a very small eigenvalue λ_1, both positive and negative: the same may occur for a C which has a minimum eigenvalue positive but very small. Thus it is natural to generalize the algorithm above to matrices C which can be inverted, but with large numerical errors, occurring when $\lambda_1/\lambda_N \sim \varepsilon \sim 10^{-16}$, the round-off level.

9.3 Constrained differential corrections [10.1]

If the normal matrix $C(\mathbf{x})$ is badly conditioned, with a very small eigenvalue $\lambda_1(\mathbf{x})$, we can generalize the definition of $C^*(\mathbf{x})$ as the matrix of the linear map which is zero on the eigenspace of the eigenvalue λ_1, containing vectors $\mathbf{v}_1(\mathbf{x})$ such that $C(\mathbf{x})\mathbf{v}_1(\mathbf{x}) = \lambda_1(\mathbf{x})\mathbf{v}_1(\mathbf{x})$, and is the inverse of $C(\mathbf{x})$ restricted to the linear subspace $H(\mathbf{x})$ orthogonal to $\mathbf{v}_1(\mathbf{x})$.

Then we can define a Constrained Differential Corrections iteration as $\mathbf{x}_{j+1} = \mathbf{x}_j + C^*(\mathbf{x}_j)D(\mathbf{x}_j)$. If the process converges, $\lim_{j\to\infty}\mathbf{x}_j = \mathbf{x}^\circ$ is such that $D(\mathbf{x}^\circ)$ is orthogonal to $H(\mathbf{x}^\circ)$ and parallel to $\mathbf{v}_1(\mathbf{x}^\circ)$. The line of points $\mathbf{x}$ satisfying this parallelism we also call LOV.

The LOV points $\mathbf{x}^\circ$ can be approximately described as local minima of the target function $Q(\mathbf{x})$ restricted to the hyper-plane $H(\mathbf{x}^\circ)$ (more exactly, they are minima of some quadratic approximation). The descent to the stream line – the LOV– along the steep sides of the mountains is performed without sliding along the stream, even if the stream is flowing, that is the function $Q(\mathbf{x})$ slowly decreases along the LOV going towards the nominal solution.

Starting from an initial guess $\mathbf{x}_0$ we can find a LOV point as the limit $\mathbf{x}^\circ$. However, this is not an effective algorithm to find the entire LOV.

10.1 The weak direction vector field [10.1]

To find an algorithm to sample a significant segment of the LOV we may use the **Weak Direction Vector field**:

$$\mathbf{F}(\mathbf{x}) = k_1(\mathbf{x})\,\mathbf{v}_1(\mathbf{x})$$

with $k_1(\mathbf{x}) = 1/\sqrt{\lambda_1(\mathbf{x})}$, is a vector field. The unit eigenvector $\mathbf{v}_1$ is not unique, $-\mathbf{v}_1$ is also a unit eigenvector. Thus $k_1(\mathbf{x})\,\mathbf{v}_1(\mathbf{x})$ is an axial vector, with well defined length and direction but an arbitrary sign. However, given an axial vector field we can define a true vector field $\mathbf{F}(\mathbf{x})$ such that the function $\mathbf{x} \mapsto \mathbf{F}(\mathbf{x})$ is continuous.

Given the vector field $\mathbf{F}(\mathbf{x})$ defined above, the differential equation

$$\frac{d\mathbf{x}}{d\sigma} = \mathbf{F}(\mathbf{x})$$

has a unique solution for each initial condition, because the vector field is smooth. If a nominal solution $\mathbf{x}^*$ has been found, let us select the initial condition $\mathbf{x}(0) = \mathbf{x}^*$, that is $\sigma = 0$ corresponds to the nominal solution, and let us denote with $\mathbf{x}(\sigma)$ the unique solution with such initial value. In the linear approximation, the solution $\mathbf{x}(\sigma)$ is one tip of the major axis of the confidence ellipsoid $Z_L(\sigma)$. Without approximations $\mathbf{x}(\sigma)$ is indeed curved and can be computed by numerical integration of the differential equation.

10.2 Sampling and parameterizing the LOV [10.1]

We had originally hoped that the solution of the weak direction differential equation and the LOV defined by the parallelism condition where the same. Later G. Tommei proved the opposite: the LOV and the solution of the differential equation are not the same, unless they are both straight lines[*]

An algorithm to compute the LOV by **continuation** from one of its points $\mathbf{x}$ is the following. The vector field $\mathbf{F}(\mathbf{x})$, deduced from the weak direction vector field $k_1(\mathbf{x})\mathbf{v}_1(\mathbf{x})$, is orthogonal to $H(\mathbf{x})$. A step in the direction of $\mathbf{F}(\mathbf{x})$, such as an Euler step for the differential equation $d\mathbf{x}/d\sigma = \mathbf{F}(\mathbf{x})$, that is $\mathbf{x}' = \mathbf{x} + \delta\sigma\,\mathbf{F}(\mathbf{x})$, is not another point on the LOV, unless the LOV itself is a straight line.

However, $\mathbf{x}'$ will be close to another point $\mathbf{x}''$ on the LOV, which can be obtained by applying constrained differential corrections, starting from $\mathbf{x}'$ and iterating until convergence at a point $\mathbf{x}^\circ$, which is on the LOV.

If the LOV parameter of the starting point $\mathbf{x}$ is σ_0, we can set $\mathbf{x}^\circ = \mathbf{x}(\sigma_0 + \delta\sigma)$, approximating a smooth parameterization of the LOV we do not know.

[*]This is an approximate statement; the difference between Differential Corrections and Newton's method also play a role.

10.3 The mixed iteration scheme for computing the LOV [10.1]

Multiple solutions, 2 steps

Convergence of multiple solutions
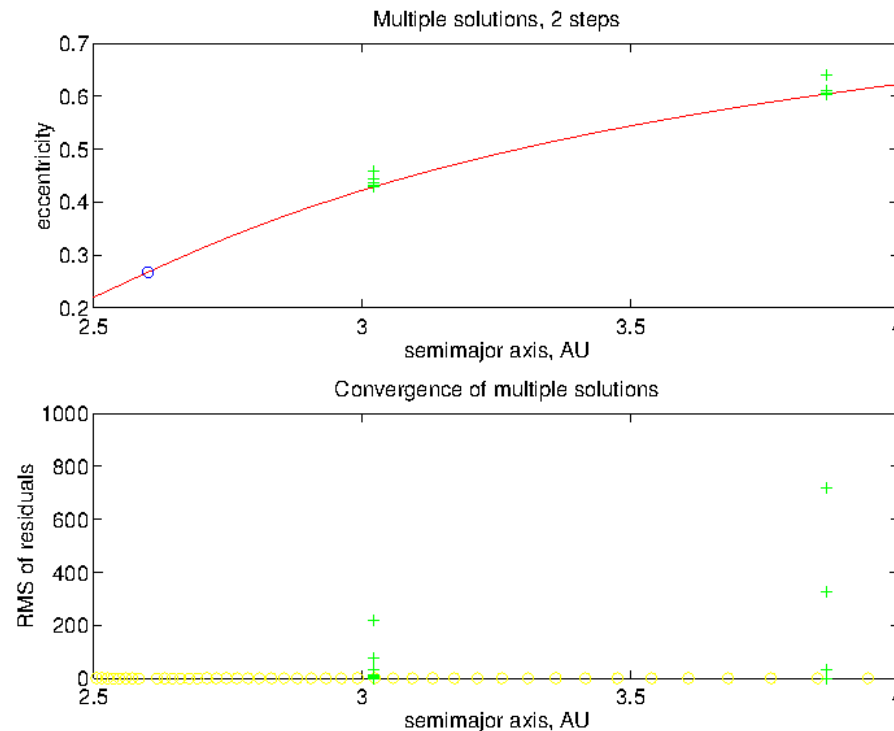
The procedure to obtain multiple solutions along the LOV; only two such solutions are shown on the (a, e) plane. Top: starting from $\mathbf{x}^*$ (circle), the LOV solutions are obtained by propagation of a solution of the weak direction equation, followed by constrained differential corrections (each iteration a green cross); they converge to the "stream" (red line), whose points have been computed by the same procedure with a much smaller step. Bottom: the RMS of the residuals is large at the starting point of constrained differential corrections, and rapidly converges towards the much smaller values obtained along the "stream" line (yellow circles).

11.1 Non-invariance of the LOV by coordinate changes [10.3]

The eigenvalues λ_j of the normal matrix C are not invariant under a coordinate change. Thus a different weak direction and a different LOV would be obtained by using some other coordinates $\mathbf{y} = \mathbf{y}(\mathbf{x})$.

This is true even when the coordinate change is linear $\mathbf{y} = S\,\mathbf{x}$: the normal matrix is transformed as $C_{\mathbf{y}} = \left[S^{-1}\right]^T C_{\mathbf{x}}\,S^{-1}$ and the eigenvalues need to be the same only if $S^{-1} = S^T$, that is if the change of coordinates is isometric.

Otherwise, the eigenvalues in the $\mathbf{y}$ space are not the same, and the eigenvectors are not the image by S of the eigenvectors in the $\mathbf{x}$ space. Thus **the weak direction and the LOV in the $\mathbf{y}$ space do not correspond** by S^{-1} to the weak direction and the LOV in the $\mathbf{x}$ space. A special case is scaling, a transformation changing the units along each axis, represented by a diagonal matrix S.

If the coordinate change is nonlinear, the same argument applies with $S = \partial \mathbf{y}/\partial \mathbf{x}$.

11.2 Examples of coordinates for initial conditions [10.3]

A non-exhaustive list of coordinates used in orbit determination is:

- Cartesian heliocentric coordinates (position, velocity)

- Cometary elements $(q, e, I, \Omega, \omega, t_p$, with t_p the time of perihelion passage)

- Keplerian elements $(a, e, I, \Omega, \omega, \ell$, with ℓ the mean anomaly)

- Equinoctial elements $(a, h = e\sin(\varpi), k = e\cos(\varpi), p = \tan(I/2)\sin(\Omega), q = \tan(I/2)\cos(\Omega), \lambda = \ell + \varpi$, with $\varpi = \Omega + \omega)$

- Attributable elements $(\alpha, \delta, \dot{\alpha}, \dot{\delta}, \rho, \dot{\rho})$, where α, δ are angular variables as seen from the observer.

For comparison, we are showing in $(\rho, \dot{\rho})$, that is (range, range-rate) plane, for 5 different coordinates, the LOV as well as the Second LOV defined by using the same procedure with the eigenvalue λ_2 and its eigenvector $\mathbf{v}_2$.

11.3 Different LOVs, for long arc [10.3]
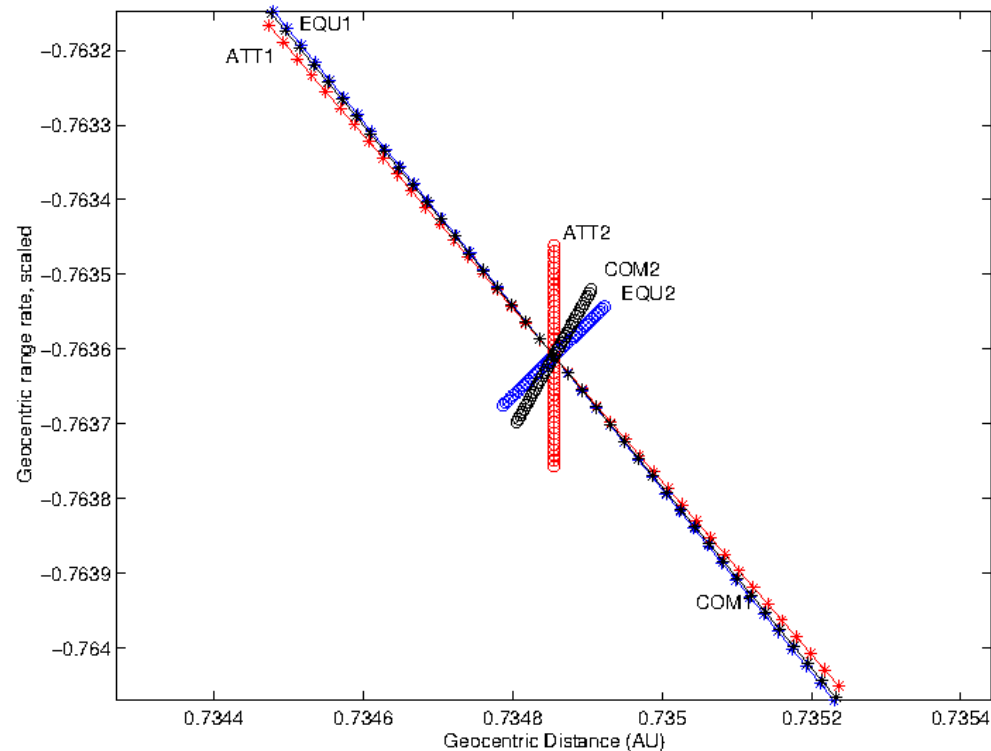
For the asteroid 2002 NT$_7$ the computation of the LOV, by using the 113 observations of the first 9 observed nights, in different coordinates. The Cartesian and attributable LOVs are indistinguishable. Anyway the different LOVs are very close, and well separated in direction from from the Second LOVs.

11.4 Different LOVs, for short arc [10.3]
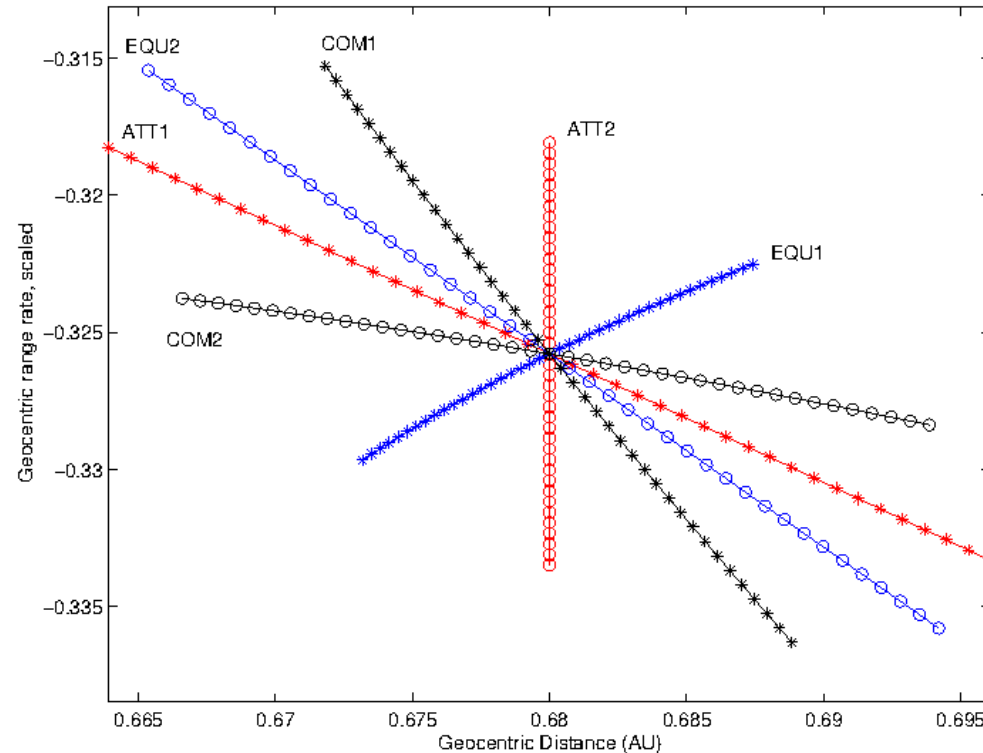
For the asteroid 2004 FU$_4$ the computation of the LOV, by using only 17 observations in the first 3 observed nights, in different coordinates; the label denotes the coordinate system used and whether the line is either the ordinary LOV or the second LOV. The Cartesian and attributable LOVs are indistinguishable on this plot and so only the attributable LOV is depicted. The other coordinates give very discordant results.

Plan Lecture 3: Line Of Variations method for impact monitoring.

In Impact Monitoring we need to find the right balance between a mathematically rigorous approach and the use of engineering safety factors. The difficulty is in that the moral responsibility we would incur in case of mistake could be enormous.

12. **LOV trace on target Planes**

13. **Computation of Minimum Distance along the LOV**

14. **Computation of Impact Probability**

15. **Operational Impact Monitoring**

12.1 The LOV trace on target planes [12.3]

A virtual impactor (VI), for a given known asteroid, is a connected set of initial conditions leading to an impact (at about the same date). Impact Monitoring (IM) is a procedure by which the VI compatible with the available observations (that is, the ones with a significant IP) are identified.

In practice, IM is possible only for a given finite time span, e.g., 100 years, and for IP above a given generic completeness limit in IP, e.g., 4×10^7 (these two values are currently used by the operational NEODyS IM system).

The method currently used by both NEODyS (CLOMON2 software) and JPL (SENTRY software) for the operational IM is based on LOV searches for impactors.

12.2 The propagation of the LOV [12.3]

In practice we select a finite segment of the LOV, truncated at some σ value, and sample it with a finite number of points (a few thousands), which we call Virtual Asteroids (VA).

Given the LOV sampling we propagate by numerical integration each VA for some finite time, e.g., 100 years; with the orbit, we propagate also the Equation of Variations. For each VA, at each close approach to Earth, we project on the TP a trace point $\mathbf{y}$, with its partial derivatives with respect to the initial conditions.

To actually assess the completeness of this search for VI, we need to take into account that of the Target Plane we can only use a finite disk $K = \{\mathbf{y} | |\mathbf{y}| < D\}$. A practical value for the maximum TP distance is in the range $0.05 \leq D \leq 0.2$ au.

12.3 Target Plane Trace of the LOV [12.4]
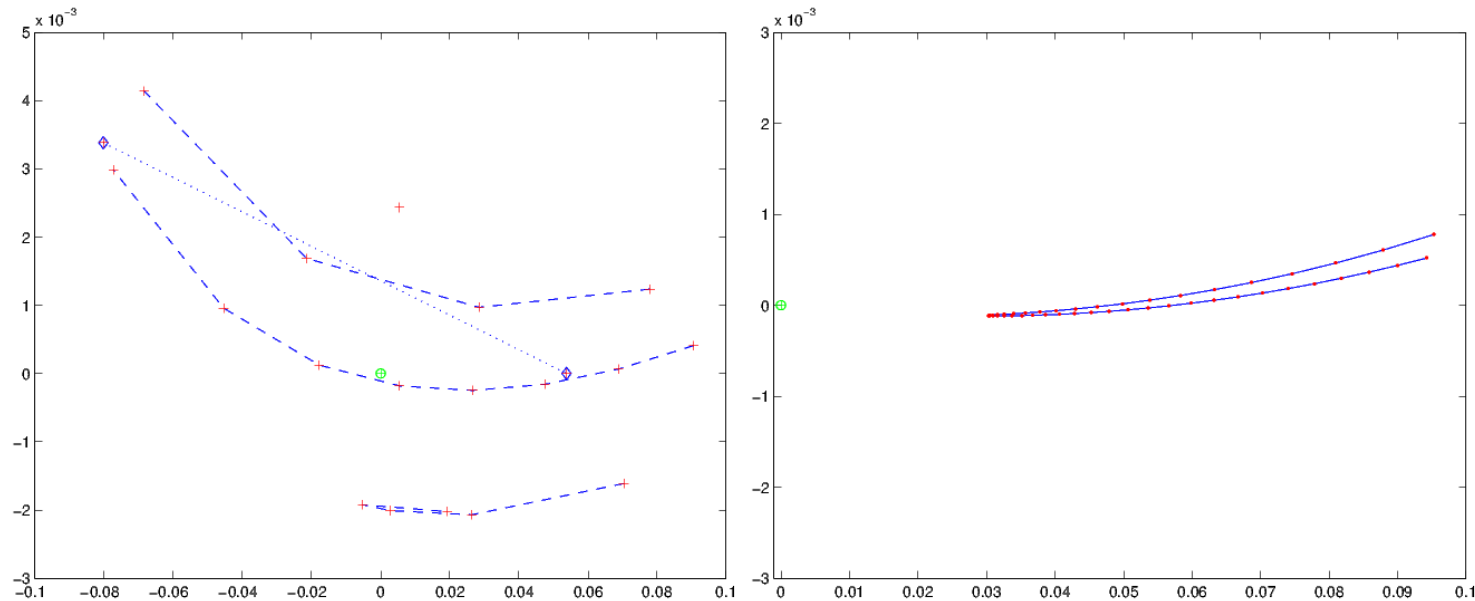
The points joined are from consecutive VAs on the LOV: they form a return. The close approaches to the same planet at about the same time form a shower, which can contain several returns. Linear interpolation is no good, as shown by return with 2 points, which actually includes a VI.

In the right figure, a new phenomenon, an interrupted resonant return: there is a resonant return in 2032 from the close approach to Earth in 2027. The asteroid has done about 3 revolutions while the Earth has done 5, but the change in a resulting from the 2027 encounter is not enough to reach the resonance value.

12.4 Detectabiliy of VIs [12.5]
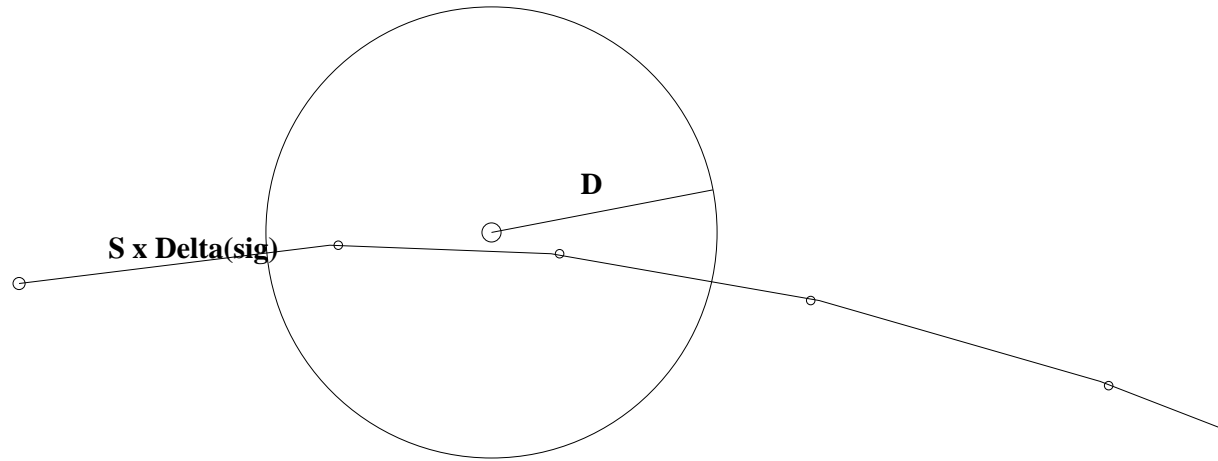
When sampling the LOV, the distance on the Target Plane between two consecutive sample points is $S \times \Delta(\sigma)$, with the Stretching $S = |\frac{d\mathbf{y}}{d\sigma}|$ and $\Delta(\sigma)$ the spacing in the LOV parameter of two consecutive sample points.

If we want two sample points to be within the disk of radius D, then $S < D/\Delta(\sigma)$.

If the LOV is sampled by $\sim 2,401$ points for $-3 \leq \sigma \leq 3$, then $\Delta(\sigma) = 1/400$ and for $D = 0.2$ au $= 4716\, R_\oplus$, then $S \leq S_{max} \simeq 1.9 \times 10^6\, R_\oplus$. By using the maximum PDF$= 1/\sqrt{2\pi}$ 9at the nominal) this corresponds to a generic completeness limit in IP for having 2 points on $K \subset TP$:

$$IP_{min} \simeq \frac{PDF(\sigma) \cdot 2b_\oplus}{S} < \frac{2b_\oplus}{S_{max}\sqrt{2\pi}} \simeq 4 \times 10^{-7}.$$

13.1 Finding the minimum possible distance [12.4]

Up to now we are just doing the same as in a Monte Carlo, just using a geometric structure to control the sampling.

Next we exploit the structure of differentiable manifold of the LOV, which potentially exists for every value of the its parameter σ. Thus the TP Trace of the LOV is a differentiable curve $\mathbf{y}(\sigma)$, with tangent vector $\frac{d\mathbf{y}}{d\sigma}$.

We find minimum possible close approach distance, at a given date, for a given LOV segment (e.g., between two consecutive VAs). The distance from the center of the Earth on the TP is $b = \sqrt{\mathbf{y} \cdot \mathbf{y}}$, the minimum occurs for $f(\sigma) = \frac{db^2}{d\sigma} = 2\,\mathbf{y} \cdot \frac{d\mathbf{y}}{d\sigma} = 0$.

The zeros of the smooth function $f(\sigma)$ can be found by any iterative scheme, such as Regula Falsi or Newton, by interpolating on the LOV in the initial state space. To start this, we just need to find two consecutive points, the one with lowest σ having $f(sigma)\frac{db^2}{d\sigma} < 0$, the other with $f(\sigma)\frac{db^2}{d\sigma} < 0$.

Then, if the function $f(\sigma)$ is defined and continuous on the interval, at least one point of minimum distance must exist and the Regula Falsi (RF) iteration (drawing straight lines between the last two points with opposite sign found) must converge to one of these minima. The usage of Newton's method has no guaranteed convergence, but allows to solve some tricky cases in which RF does not apply, including those with just one VA in the disk K.

13.2 Complexity of the TP trace of the LOV [12.4]
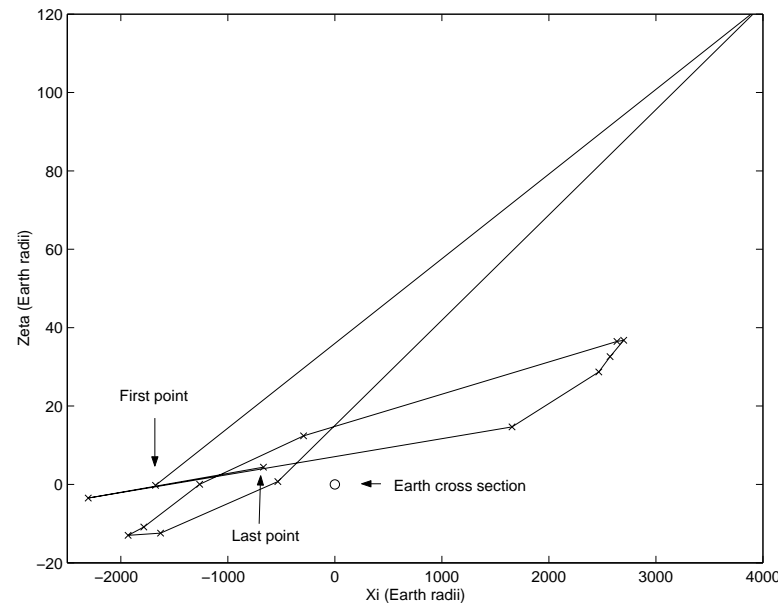
A sequence of consecutive VA (intersecting the TP around a given date) corresponds to a continuous set, a segment of the LOV. Intermediate points can be interpolated, in two steps: following the weak direction for $\delta\sigma < \Delta\sigma$, then going down to the LOV by constrained differential corrections.

However, this may not give curve segment of TP trace points joining the first and the last point: the intermediate points could be outside of the disk K. Even if this happens, the behavior of the segment inside K could be wild, as in the figure. A RF starting from the first and the last point would converge, but to a local minimum of the distance, not the absolute minimum, and a VI could be missed.

13.3 The principle of the simplest geometry [12.4]

The behavior of the TP trace between two consecutive VA could be made less wild by denser sampling of the LOV. However, this would increase the computational load, unless a smart adaptive system of densification is used (work in progress).

The alternative is to use the argument that wild behaviors, as in the last figure, require an increase of the stretching to a value much larger than $|\mathbf{y}(\sigma + \Delta\sigma) - \mathbf{y}(\sigma)/\Delta\sigma$. Thus the VI which could be found would have IP below the generic completeness limit.

This is why we adopt the principle of the simplest geometry, by which the curve segment does not exit the TP disk of radius d_{max}: then there needs to be at least one minimum of the closest approach distance.

Moreover, the VI with the largest IP can be found when the behavior between the two consecutive VA is simple, with a single minimum (see figure): only 2 cases, simple and interrupted return. In our software there are some provisions for identifying two local minima of the distance between two consecutive VA, but we cannot guarantee they work always.

If these two simplifying assumptions apply, then the generic completeness is an estimate of the level of IP above which the VI should be found.
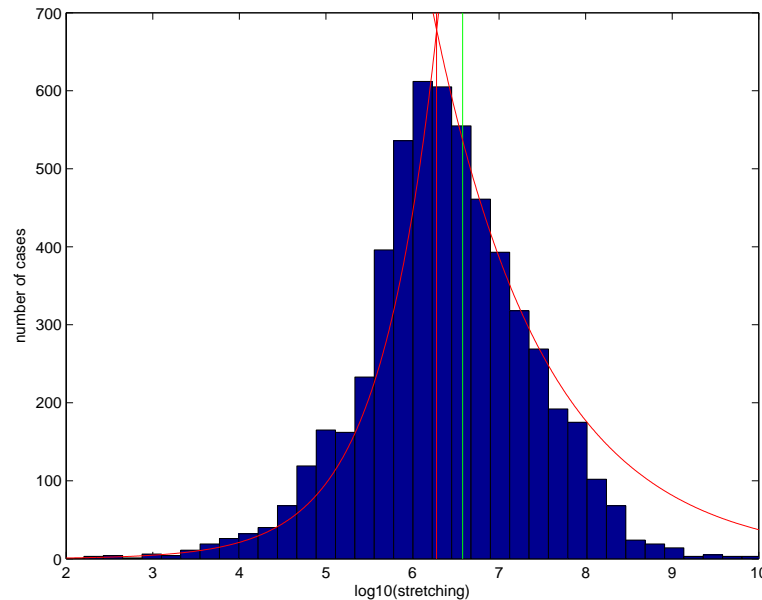
13.4 The Reliability of lists of VIs
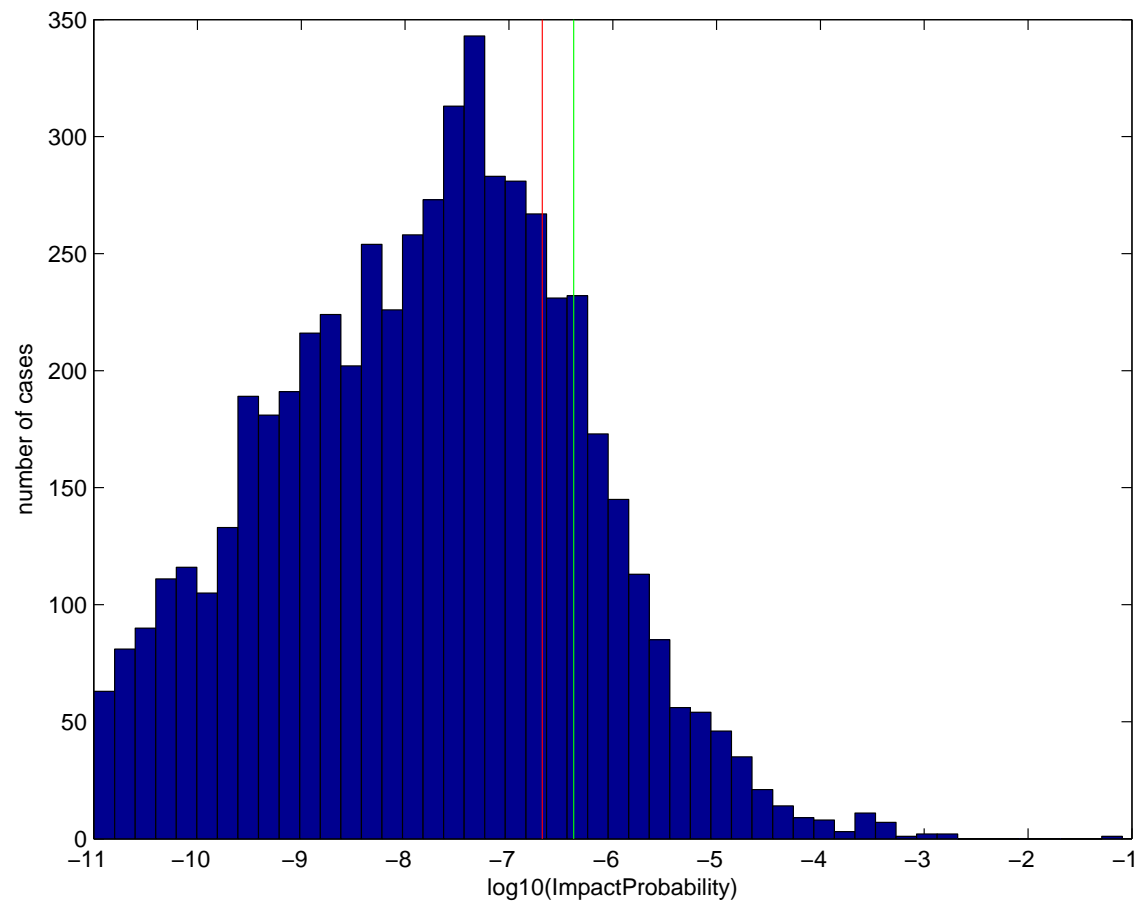
Histogram of all the VIs in the NEODyS risk files, as a function of the Impact Probability (IP), on May 5, 2011.

Red: generic completeness limit for two consecutive VA in the disk of radius $D = 0.2$ au. Green: the same for one VA in the disk.

13.5 The Probabilistic Detection of VIs: Stretching

Histogram of all the VIs in the NEODyS risk files, as a function of the Stretching (in Earth radii), on May 5, 2011. The vertical lines are the generic completeness limit for two points in the disk, red, and for one point in the disk, green.

The growing exponential is $N = 0.048 \cdot S^{0.66}$. The decreasing exponential is N multiplied by the ratio $1.9 \times 10^6/S$. Thus detection of a VI with stretching in the $10^7 \div 10^8$ range is random, with a different code we can get another one or none.

The relationship $N \sim S^{2/3}$ is universal, just from the mathematics of chaos, a property of our solar system, an artifact of our discovery and computational methods?

14.1 Computation of Impact Probability: 1-dimensional

The simplest way to address the computation of the IP is to use a 1-dimensional approach, namely, by computing the probability of being inside the Earth impact cross section conditional to the LOV trace on the TP.

Let us suppose we are able to find that the LOV segment which leads to impact is parameterized by the segment $[\sigma_1, \sigma_2] \subset \mathbb{R}$. Then the IP is computed as

$$\mathcal{P}_{\Sigma}([\sigma_1, \sigma_2]) = \int_{\sigma_1}^{\sigma_2} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\sigma^2}{2}\right) d\sigma \simeq (\sigma_2 - \sigma_1) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\sigma_c^2}{2}\right)$$

where $\sigma_c \in [\sigma_1, \sigma_2]$, in an approximation applicable for a very short interval of σ. (Note that the generic completeness limit is based on this 1-dim model.)

This approximation is appropriate when the **Width**, which is the short axis of the linear confidence ellipse on the TP, is much less than $R_{\oplus}$, e.g., few km. Then the full 2-dim probability integral would give the same. Another approximation could be to use the σ coordinates of the first and the last VA belonging to the VI (that is, impacting) as approximations of σ_1, σ_2, respectively. This is a good approximation if the VA impacting are > 10.

14.2 Computation of Impact Probability: 2-dimensional

For a full 2-dimensional estimation of the IP, taking into account a finite Width W as well as a Stretching S, we compute the probability integral directly on the TP.

We assume that at the closest approach point $\mathbf{y}_m = (\xi_m, \zeta_m)$, found by RF, the LOV tangent is parallel to the ζ axis (by a suitable rotation on the TP). If $\mathbf{y}_0 = (\xi_0, \zeta_0)$ was the trace on the TP of the nominal, the linear PDF would be

$$\mathcal{P}_{\Xi,Z}(\xi,\zeta) = \frac{1}{2\pi} exp \left\{ -\frac{1}{2} \left[\left(\frac{\xi - \xi_0}{W} \right)^2 + \left(\frac{\zeta - \zeta_0}{S} \right)^2 \right] \right\},$$

but his is inapplicable: the point $\mathbf{y}_0$ may not even exist (i.e., the nominal might not have that close approach). If σ_m is the value of the LOV parameter corresponding to $\mathbf{y}_m$, the PDF at the same point is

$$\mathcal{P}_{\Xi,Z}(\xi_m,\zeta_m) = \frac{1}{2\pi} exp \left[-\frac{1}{2} \sigma_m^2 \right].$$

Then an approximation of the PDF near the point $\mathbf{y}_m$ can be:

$$\mathcal{P}_{\Xi,Z}(\xi,\zeta) = \frac{1}{2\pi} exp \left\{ -\frac{1}{2} \left[\left(\frac{\xi - \xi_m}{W} \right)^2 + \left(\frac{\zeta - \zeta_m}{S} + \sigma_m \right)^2 \right] \right\}$$

to be used to compute numerically the probability integral over the impact cross section $\{|\mathbf{y}| \leq B\}$.

14.3 Computation of Impact Probability: Off-LOV VI

The previous method to compute the IP is robust with respect to non-linearities along the LOV trace, because it uses the PDF value at a nearby point. However, it assumes linearity in the transversal direction, thus it may fail if W is large.

This can be problem if the minimum close approach distance along the LOV is $|\mathbf{y}_m| > B$, that is the LOV point $\mathbf{y}_m$ does not correspond to an impact. If W is large, $W >> B$, the PDF computed by the previous approximation may be small but not zero over the impact cross section. However, the IP computation is not robust with respect to non-linearities in the direction orthogonal to the LOV trace.

The Impact Monitoring systems NEODyS (in Pisa) and SENTRY (at JPL) use diferent approaches to handle this problem. JPL uses linearity tests at convergence of the RF iterations. NEODyS applies the principle that an explicit **VI representative**, an intial condition leading to an impact, needs to be found for each VI. This is searched by Newton's method on the TP, combined with the construction of the regression subspace of $\mathbf{x}$ given $\mathbf{y}$ on the TP (similar to the semilinear method).

This procedure has a number of difficulties, including the possible non-convergence of Newton's method and the need to correct somehow the IP computation. A fully satisfactory approach has not yet been found for these off-LOV VI cases. Fortunately, these cases must have very low IP.

15.1 Operational Impact Monitoring

The **NEODyS online information system** (http://newton.dm.unipi.it/neodys) provides astrometric and orbital information on all Near Earth Asteroids (NEA). It includes a **Risk List** with all the known NEA which can impact the Earth in the next 100 years (plus some later). Currently (5/9/2014) the risk list contains 453 NEA.

The information from NEODyS are also disseminated by ESA **Near Earth Objects Coordination Center** at ESRIN, Tor Vergata. In the future (2016, TBC) NEODyS will be migrated to ESRIN and managed under ESA responsibility. Since 2002 the NEODyS service is fully duplicated by NASA JPL (http://neo.jpl.nasa.gov/risk/).

15.2 What is in a Risk File: example 1

Object 2010 AR$_{85}$

date	dist	+/-	width	stretch	IP	exp. en.	PS
YYYY/MM	(RE)		(RE)	RE/sig		MT	
2012/02/04.288	3138.83	+/-	1653.	1.05E+05	1.16E-10	4.12E-05	-4.29
2013/02/05.841	1300.70	+/-	1365.	1.41E+05	1.53E-09	4.50E-04	-3.41
2014/02/07.369	417.99	+/-	1234.	1.77E+05	2.36E-09	6.22E-04	-3.38
2015/02/08.447	99.19	+/-	1159.	2.14E+05	2.20E-09	5.41E-04	-3.53
2016/02/09.260	442.69	+/-	1111.	2.52E+05	1.77E-09	4.14E-04	-3.72
2017/02/08.926	685.16	+/-	1077.	2.90E+05	1.37E-09	3.08E-04	-3.91
2018/02/09.493	867.36	+/-	1052.	3.27E+05	1.05E-09	2.31E-04	-4.09
....							

The Width approximates the RMS uncertainty in the MOID. The Stretching is moderate, the Width is very large: it can have MOID > 0.05 (not even be a PHA).

The Expected Energy divided by the small IP is the Energy the impact would have, if it was to occur. In this case the energy is about $300,000$ MegaTons, because the asteroid has $H = 17.4$, that is diameter between 1.0 and 2.2 km. This was written in early 2011; later this asteroid was identified with 2011 WS$_2$, not a PHA (MOID$= 0.09$ au) and does not approach closer than 0.12 au in this century.

15.3 What is in a Risk File: example 2

Object 2010 RF$_{12}$

date YYYY/MM	dist (RE)	+/-	width (RE)	stretch RE/sig	IP	exp. en. MT	PS
2095/09/05.993	1.06	+/-	0.000	2.05E+01	8.08E-02	7.30E-04	-3.13
2096/09/04.914	0.97	+/-	0.000	1.01E+04	1.02E-04	9.21E-07	-6.04
2098/09/05.355	1.53	+/-	0.000	2.65E+05	6.16E-06	5.56E-08	-7.27
2098/09/05.244	1.55	+/-	0.000	2.32E+05	5.22E-06	4.72E-08	-7.34
2099/09/05.950	2.25	+/-	0.000	3.45E+05	4.52E-06	4.10E-08	-7.41
2099/09/05.560	1.88	+/-	0.002	1.42E+05	8.01E-06	7.25E-08	-7.16
2100/09/05.789	1.56	+/-	0.000	3.53E+05	4.73E-06	4.27E-08	-7.39
2100/09/05.801	0.70	+/-	0.000	7.85E+04	9.73E-06	8.79E-08	-7.08

The Expected Energy is much less than the IP, indeed the impact energy would be 9 KiloTons (damage on the ground very unlikely). The IP is large: 0.08.

The Stretching is similar to the other case, but this is after 85 years, not 2-3. This is a direct impact, the Stretching decreases between 2012 and 2095.

With observations in either 2011 or 2012 (but apparent magnitude 26) this object should have a IP for 2095 either 0 or 1. This was written in 2011; actually, it has not been reobserved. It could be recovered in 2047.

Plan Lecture 4: Experimental Impact Monitoring

An operational Impact Monitoring is a commitment from which it is very difficult to withdraw/retire.

16. **Well determined orbits and scattering plane**

17. **Yarkovsky effect and Impact Monitoring**

18. **The most extreme example: Bennu**

19. **An endless job**

16.1 The well constrained case

As an asteroid is re-observed after the discovery, the orbit quality improves. Sharp transition to a well constrained orbit as soon as a second apparition is observed.

The confidence region, even propagated for decades, projects into a small ellipsoid in space, with a major axis a small fraction of an au. If this happens, even assuming the **Minimum Orbit Intersection Distance (MOID)** is small (there are two points on the asteroid's and on the Earth's osculating ellipses at a short distance), the asteroid is unlikely to be near the MOID point when the Earth is there.

Anyway, if there was a VI, it would be detectable by linear propagation of the PDF. Thus multiapparition, and even more numbered, asteroids do not appear in the output of our automated Impact Monitoring systems.

However, the condition of very low MOID can last for long times ($50 \div 100$ years typical), being controlled by secular perturbations in e, ω. What if there is a very close approach (not a collision) within this time interval?

16.2 The scattering encounter

Example: (9942) Apophis has a very deep encounter with Earth on 13/4/2029, at $38\,000$ km from geocenter.

The runoff (long semi-axis of confidence ellipsoid) increases at once by a factor $\sim 4\,300$; by 2037 propagation increases this value to $40\,000$ times the value before 2029. Thus the width of keyhole (for definition, cfr. Valsecchi lecture) for a resonant return with impact either in 2036, with 6/7 resonance, or in 2037 with 7/8 resonance is $\sim 2b_\oplus/40\,000 \sim 600$ m.

From then on, the runoff increases exponentially with time: the average time span between close approaches is a local (in time) version of the Lyapounov time. By 2150 the runoff is ~ 100 au.

Essentially, after 2029 we are back to the case of a poorly determined orbit, with the difference that the large uncertainty in the asteroid state is due to divergence of nearby orbits, not to original poor constraints of the initial conditions.

A very close approach like the one of Apophis in 2029 is very rare (for asteroids of comparable size). On 13/4/2029 Apophis will be visible to naked eye.

16.3 Exponential divergence due to close approaches

For 2009 FD, the same phenomenon occurs over a longer time span.
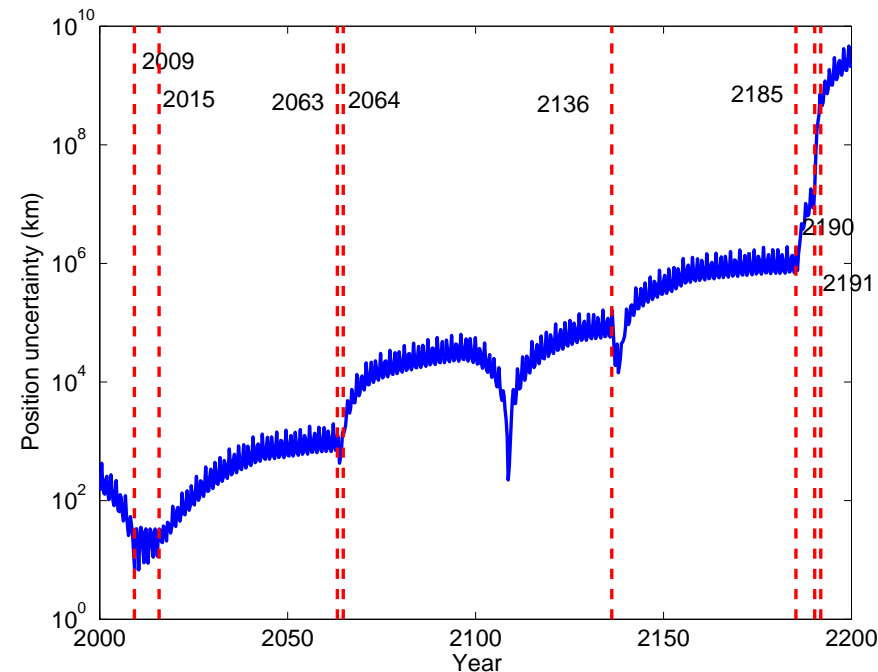
The divergence of nearby orbits, measured by the runoff. A sequence of close approaches, especially the "double hits" in 2063/2064 and again in 2136/2137, increase the uncertainty of the orbit increases by a factor $\sim 3 \times 10^5$ between 2009 and 2185, corresponding to a Lyapounov time of ~ 14 years.

The MOID is not ~ 0 after 2200: with only 4 node crossing intervals in a period of ω (about $16\,600$ years), the Lyapounov time estimate increases. The rigorous Lyapounov exponent, the limit for $t \to +\infty$, may well be zero (Marchal'a conjecture).

16.4 The scattering encounter metric for LOV

If there is a scattering deep encounter the IM problem is best studied by the trace of the VA swarm on the TP of the encounter: 2029 for Apophis, 2135 for Bennu, 2185 for 2009 FD.

The trace on the scattering TP can be studied by both Monte-Carlo and LOV methods: the VA form a line, or a very narrow strip. The line can be parameterized by some parameter, like σ for the LOV and χ for the Monte Carlo.

The LOV method is more efficient computationally, but the TP trace might turn out to be shorter because of the choice of the LOV, which depends upon the metric used in the orbital elements space.

A recent improvement has been to use a scattering plane (semi)metric, obtained by propagating the nominal orbit to the scattering TP and by pulling back the TP natural metric. This LOV is efficient and does not risk losing some possible VI near the ends. This technique has been used in the Apophis example which follows.

17.1 Yarkovsky effect and IM

If the orbit is already well determined, can small dynamical effects affect IM?

The Yarkovsky effect is a non-gravitational secular perturbation, due to anisotropic thermal emission. To assess its relevance, an order of magnitude estimate:

A typical value for the secular $da/dt = 10^{-9}$ au/y for a ~ 1 km diameter Near Earth Asteroid with $a \simeq 2$ au.

Computation of along track acceleration: $a\Delta\lambda \simeq -\frac{3}{4}\Delta a \left(\lambda - \lambda_0\right)$

Example: if $da/dt = 10^{-9}$ au/y, $\Delta t = 30$ y, $(\lambda - \lambda_0) = 63$ rad, that is 10 periods, then $a\Delta\lambda \simeq 1.4 \times 10^{-6}$ au $\simeq 1/30 R_\oplus$. Even after $\Delta t = 100$ y it is still $\sim 1/3\, R_\oplus$. It may move the VI along the LOV, neither create nor destroy a VI.

Conclusion: if there is no intermediate encounter with a planet, the Yarkovsky effect is not enough to change the IM result. However, if there is amplification, such as the one resulting from a deep encounter and/or multiple encounters, then the change due to Yarkovsky can be much larger than the keyhole width.

17.2 Measuring the Yarkovsky effect

The Yarkovsky effect can be determined in two ways:

1. an empirical parameter, transversal acceleration, is estimated together with initial conditions. Then the least squares fit to observations gives a solution for 7 variables, with a covariance matrix 7×7;

2. physical parameters obliquity, density and thermal conductivity, are estimated by a combination of photometric, radar and infrared observations. Then the secular effect can be computed by a model including a solution of the heat equation, following Farinella (1984), Vokrouhlickćy et al. (2000), and later nonlinear models.

As for approach 1, there are now 21 asteroids with Yarkovsky effect measured from 7 parameter Orbit Determination (S/N$> 3\sigma$), plus 13 (S/N$> 2\sigma$), all of them NEA (Farnocchia et al. Icarus 2013)

As for approach 2, the problem is there is no case in which all the relevant parameters are measured (this would require a space mission such as Osiris/Marco Polo). Thus the uncertainty of the non-measured quantities needs to be represented by a statistical model.

17.3 Apophis: models for Yarkovsky
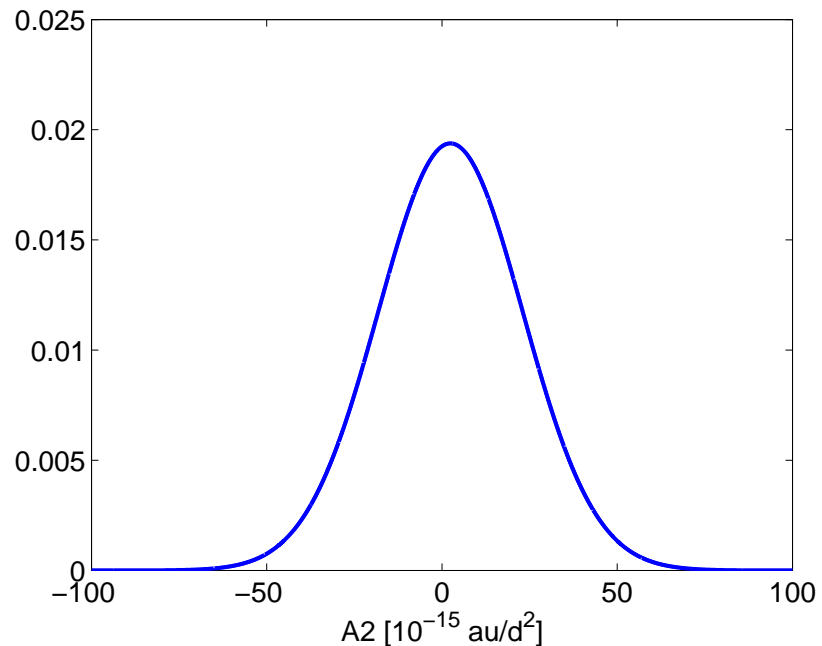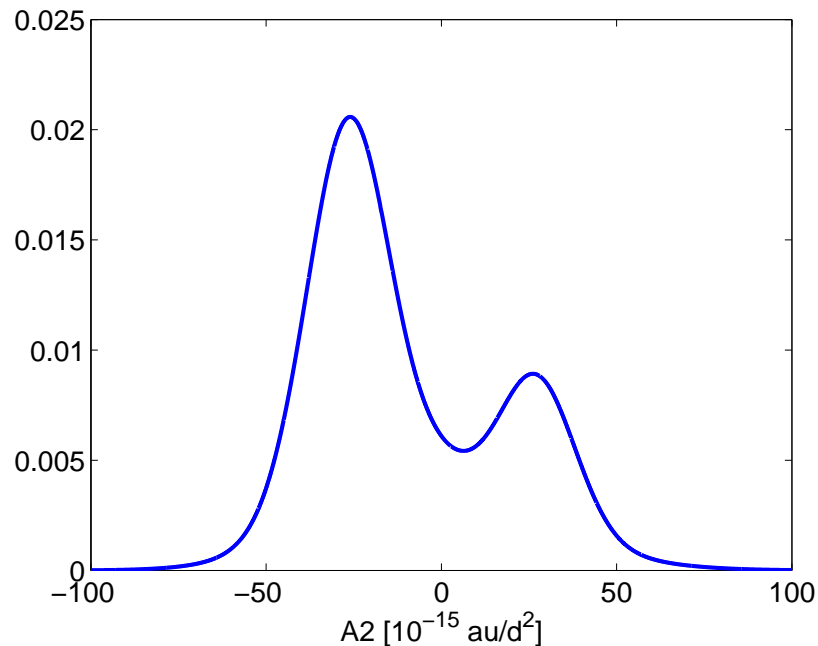
Yarkovsky effect model with a method from Farnocchia et al. (Icarus 2013). For each model parameter, a PDF is computed, e.g., diameter and albedo from Herschel, density from values for same composition and size, frequency of direct vs. retrograde from population models, thermal inertia from Delbò et al. 2007. A Monte Carlo composes a PDF for the transversal thermal effect averaged A_2 (left figure).

With the use of the 2013 radar measurements, the Yarkovsky effect is weakly determined by the 7-parameters orbit determination (right figure).

17.4 Apophis: scattering plane map
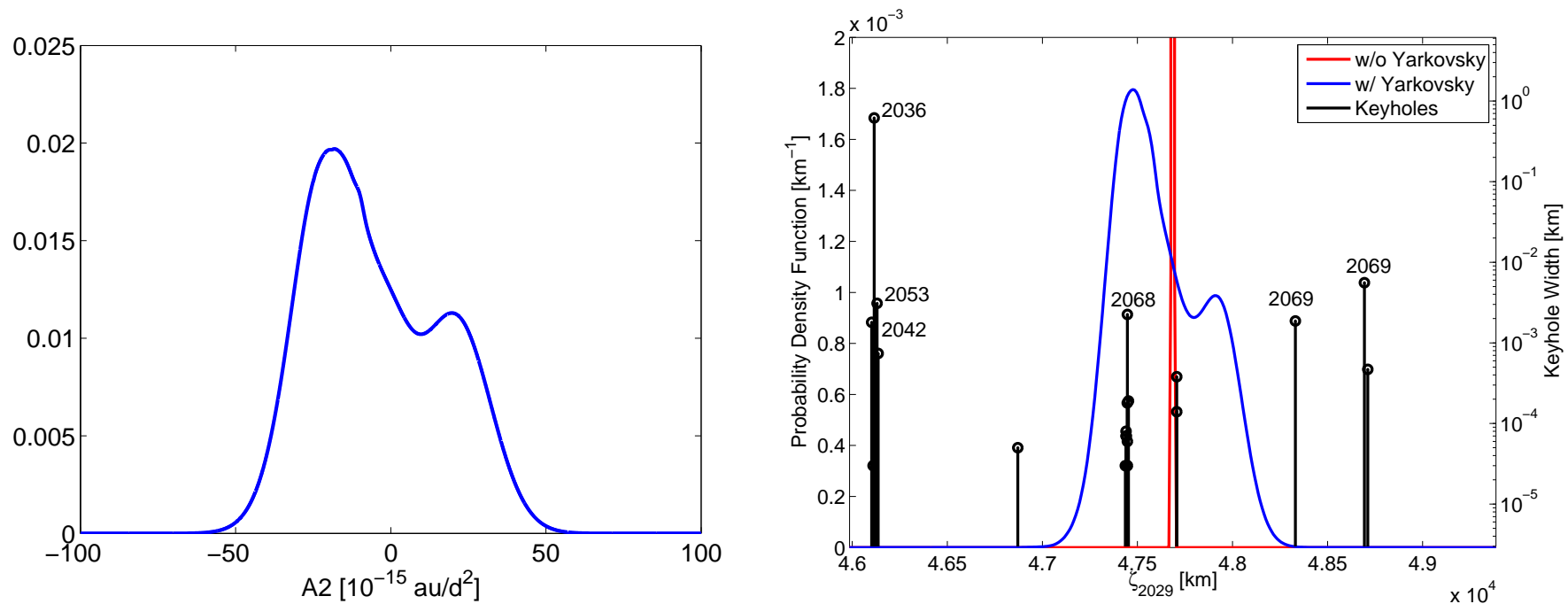
The two PDF from previous page can be combined by multiplication, which is the right Bayesian approach taking into account both sources of information: Yarkovsky model and orbit determination (left figure).

The trace on the 2029 TP is computed and the PDF propagated to it (blue line). The right figure overlaps vertical bars representing keyholes, with altitude indicating width. The the IP associated to each keyhole is just the product PDF $\times$ width.

Red line: illusory estimate of TP trace for zero Yarkovsky (STD just 3 km).

17.5 Risk files for Apophis

Question (ESA): how reliable are the detections of VIs and the IP computed? We cannot give a standard deviation of the IP! We can test for model dependencies. Two computations as different as possible, done in Pisa (Milani-Spoto, top) and at JPL (Chesley-Farnocchia, bottom)

Object: 99942, no a priori, DE405, LOV with scattering plane metric

date YYYY/MM/DD	σ_{LOV}	dist $R_\oplus$		width $R_\oplus$	stretch $R_\oplus/\sigma$	$p(R_\oplus)$	Palermo Scale
2068/04/12.634	-1.317	1.73	+/-	0.006	3.85E+5	1.88E-6	-3.67
2068/10/15.324	-0.182	1.48	+/-	0.000	5.29E+6	2.39E-7	-4.57

Object: 99942, with a priori Yarkovsky model, DE424, Monte Carlo

date YYYY/MM/DD	σ_{LOV}	dist $R_\oplus$		width $R_\oplus$	stretch $R_\oplus/\sigma$	$p(R_\oplus)$	Palermo Scale
2068/04/12.640	-0.663	0.04	+/-	0.000	6.48e+05	3.90e-06	-3.32
2068/10/15.400	0.358	1.10	+/-	0.000	1.79e+07	1.50e-07	-4.75

Discrepancy by a factor up to 2 in IP does not matter because:

1) values are small anyway;

2) there is time for deflection until the 2051 encounter, with increase by > 100 in runoff. Waiting after 2029 to confirm the 2068 impact does not matter.

17.6 1950 DA: Yarkovsky model
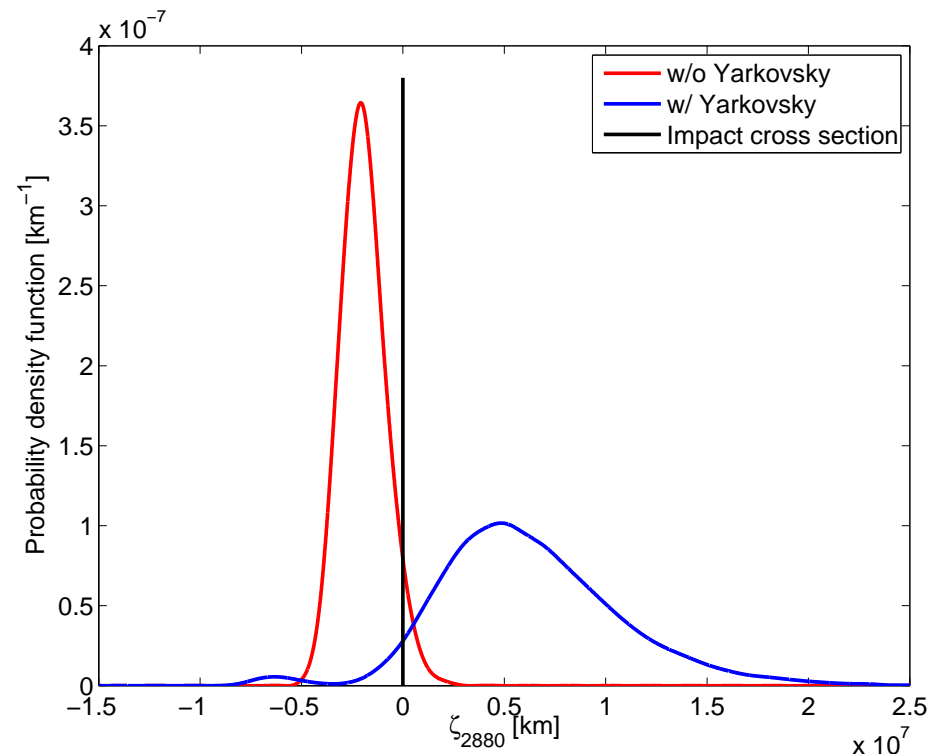
Farnocchia and Chesley (submitted 2013). IP in 2880: 0.0004. Palermo Scale $= -0.58$ (largest), because it is > 1 km diameter.

This VI is for a very remote time, but IP is 26% of the background risk for such big objects, over the timespan up to 2880. This requires some contorted argument to claim that the Spaceguard Survey sponsored by NASA has achieved a 90% reduction in the asteroid impact risk.

Alternate computation: to be done, requires to use the same DE431 ephemerides.

18.1 Bennu: best measure of Yarkovsky effect

Solving for initial conditions plus one parameter, taking into account the recent data, including 2011 radar observations:

Transversal acceleration $A_T = \frac{A_2}{r^{2.75}}$ with solution for the empirical constant $A_2 = -47.5 \pm 0.2 \times 10^{-15}$ au/d^2, corresponding to $da/dt = 1.90$ au/billion years.

For given Yarkovsky model S/N~ 200; $1\sigma \sim 0.01$ au/billion years ~ 1.5 m/y in a; transverse acceleration $1\sigma \sim 4 \times 10^{-15} m/s^2 \sim 4\ fm/s^2$ (femtometers). Incredible accuracy!

However, such accuracy requires a dinamical model accuracy which is a challenge. Problems are: 1) Yarkovsky models 2) Ephemerides DE405/DE424/DE430 3) General Relativity, including Earth 4) Perturbing asteroids BIG25/BIG16/CPVH 5) Astrometric treatment, including outlier removal.

18.2 Bennu: dynamical model problems

Dynamical Effect of Several Model Variations

Solution Model	$\delta da/dt$ $(10^{-9}$ au/y)	$\delta\zeta_{2135}$ (km)
Yarkovsky Models		
$d = 2.00$	0	0
$d = 2.75$	-0.0172	-9788
Nonlinear	0.004	28432
Asteroids		
25 Perturbers	0	0
BIG-16 only	-0.0003	-213
CPVH only	-0.0010	-3933
Relativity		
Full EIH	0	0
w/o Venus	0.0016	9638
w/o Earth	0.0289	171533
w/o Jupiter	-0.0012	-8440
Other		
Area/Mass$= 0$	-0.0005	-7292
DE405 w/BIG16	-0.0048	-15150

18.3 Bennu: scattering plane map
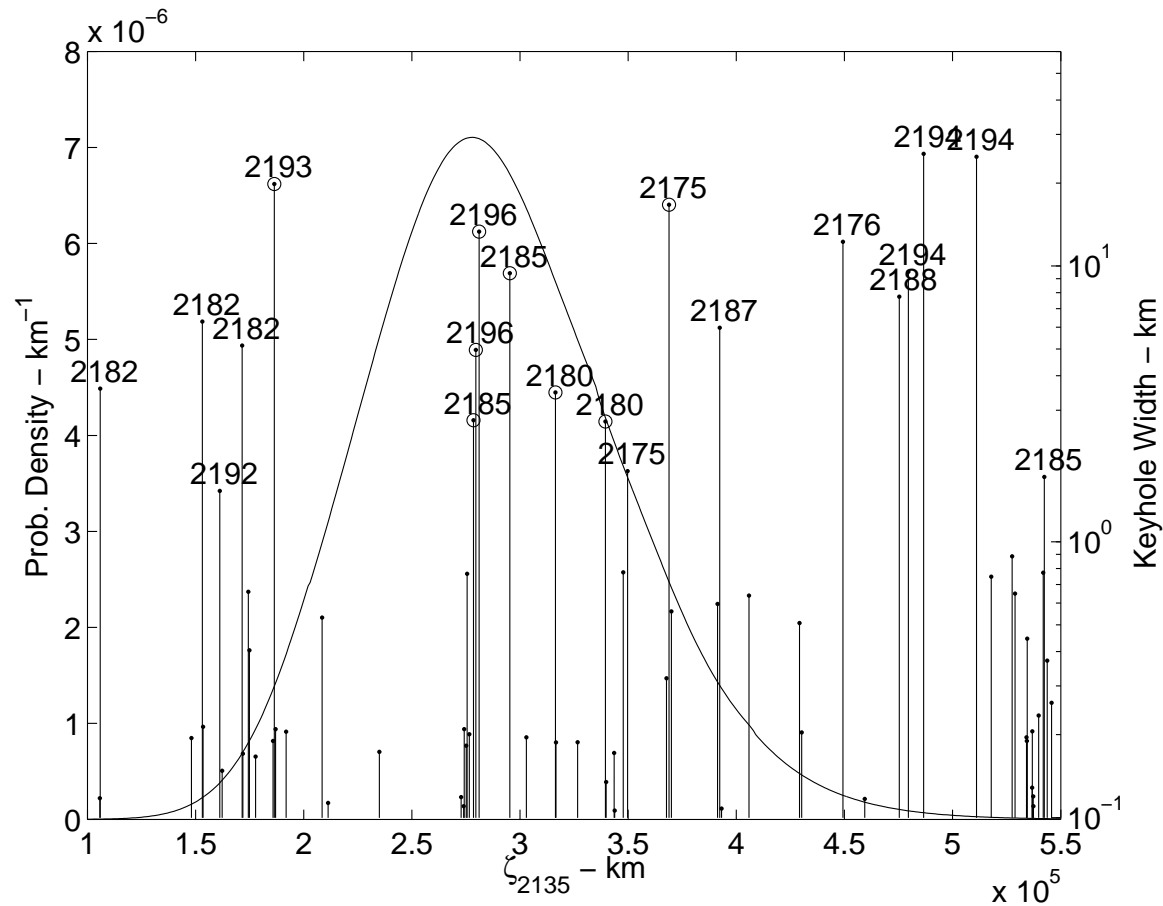
The trace on the 2135 TP is computed and the PDF propagated to it (curve). The vertical bars represent keyholes, with altitude indicating width. The the IP associated to each keyhole is just the product PDF $\times$ width.

18.4 Bennu: new risk file

Year	ζ_{2135} (km)	ζ-width (km)	Impact Prob.	Palermo Scale
2193	186415	19.84	2.75E-5	2.85
2185	278479	2.76	1.96E-5	2:98
2196	279590	4.96	3.52E-5	2:75
2196	281070	13.32	9.45E-5	2:32
2185	295318	9.42	6.33E-5	2:47
2180	316352	3.48	1.95E-5	2:96
2180	339506	2.73	1.14E-5	3:20
2175	368877	16.67	4.13E-5	2:63

A simplified risk file for the impact risk from (101955) Bennu in the 22nd century (only VIs with IP$> 10^{-5}$). ζ is the time of arrival coordinate on the 2135 TP, ζ-width is the keyhole width.

The cumulative IP is 3.7×10^{-5}, the cumulative PS is -1.70.

These results from paper by Chesley et al., Icarus 2014. This risk file has now been published on SENTRY/NEODyS.

19 Conclusion: endless job

We should, to comply with our duty of "protecting the Earth from impacts":

1. continue to take care of new discoveries

2. follow up old cases forever, pushing forward the time horizon

3. monitor the adequacy of the dynamical model and upgrade when necessary

4. determine Yarkovsky effect whenever possible, even if S/N low

5. be ready for immediate impactors