Projection Methods

Geometric Numerical Integration

Seminar WS 05/06

Contents

Projection Methods

Suppose we have an $(n-m)$-dimensional submanifold of $\mathbb{R}^n$,

$$M = \{y : g(y) = 0\}$$

$(g : \mathbb{R}^n \to \mathbb{R}^m)$, and a differential equation $\dot{y} = f(y)$ with the property that

$$y_0 \in M \quad \textit{implies} \quad y(t) \in M \textit{ for all } t.$$

The last assumption is equivalent to $g'(y)f(y) = 0$ for $y \in M$.

Definition (Weak Invariant)

We call g(y) a **weak invariant**, if $g'(y)f(y) = 0$ for $y \in M$; and we say that $\dot{y} = f(y)$ is a **differential equation on the manifold M** in the situation above.

Example (Invariant vs. Weak Invariant)

Our assumption by the definition of a weak invariant is really weaker than the requirement that all components $g_i(y)$ of $g(y)$ are invariants in the sense of an earlier definition: we only require $g'(y)f(y) = 0$ for $y \in M$ and not $g'(y)f(y) = 0$ for all $y \in \mathbb{R}^n$.

Example (Pendulum Equation)

Consider the pendulum equation written in Cartesian coordinates:

$$\dot{q}_1 = p_1, \qquad \dot{p}_1 = -q_1\lambda,$$

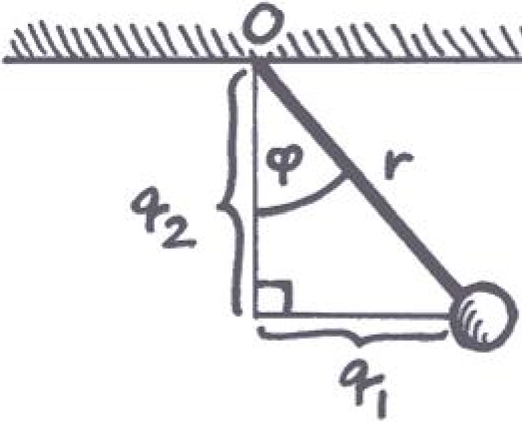
$$\dot{q}_2 = p_2, \qquad \dot{p}_2 = -1 - q_2\lambda,$$

where $\lambda = ({p_1}^2 + {p_2}^2 - q_2)/({q_1}^2 + {q_2}^2)$. (One can check by differentiation that $q_1 p_1 + q_2 p_2$ is an invariant (orthogonality of the position and velocity vectors).)

The length of the pendulum ${q_1}^2 + {q_2}^2$ is only a weak invariant.

There are methods which conserve quadratic first integrals (for example the implicit midpoint rule) but not the quadratic weak invariant ${q_1}^2 + {q_2}^2$.

No numerical method that is allowed to evaluate the vector field $f(y)$ outside M can be expected to conserve weak invariants exactly.


O
φ
r
q2
q1

$$q_1 = rsin\phi \qquad p_1 = r\dot{\phi}cos\phi$$
$$q_2 = -rcos\phi \qquad p_2 = r\dot{\phi}sin\phi$$

Compare

$$\dot{p}_1 = r\ddot{\phi}cos\phi - r\dot{\phi}^2 sin\phi$$
$$\dot{p}_2 = r\ddot{\phi}sin\phi + r\dot{\phi}^2 cos\phi$$

with

$$\dot{p}_1 = -q_1\lambda = -rsin\phi\frac{r^2\dot{\phi}^2 + rcos\phi}{r^2}$$
$$\dot{p}_2 = -1 - q_2\lambda = -1 + rcos\phi\frac{r^2\dot{\phi}^2 + rcos\phi}{r^2}$$

to get

$$r\ddot{\phi} = -sin\phi$$

Definition (Standard Projection Method)

Assume that $y_n \in M$. One step $y_n \mapsto y_{n+1}$ is defined as follows:

- Compute $\tilde{y}_{n+1} = \Phi_h(y_n)$, where Φ_h is an arbitrary one-step method applied to $\dot{y} = f(y)$;
- project the value $\tilde{y}_{n+1}$ onto the manifold M to obtain $y_{n+1} \in M$.

For $y_n \in M$ the distance of $\tilde{y}_{n+1}$ to M is of the size of the local error, i.e., $O(h^{p+1})$.
Therefore, the projection does not deteriorate the convergence order of the method.

For the computation of y_{n+1} we have to solve the **constrained minimization problem**

$$\|y_{n+1} - \tilde{y}_{n+1}\| \;\rightarrow\; min$$

subject to

$$g(y_{n+1}) = 0.$$

A standard approach is to introduce **Lagrange multipliers** $\lambda = (\lambda_1, ..., \lambda_m)^T$, and to consider the **Lagrange function**

$$L(y_{n+1}, \lambda) = \|y_{n+1} - \tilde{y}_{n+1}\|^2 /2 - g(y_{n+1})^T \lambda.$$

The necessary condition $\partial L/\partial y_{n+1} = 0$ then leads to the system

$$y_{n+1} = \tilde{y}_{n+1} + g'(\tilde{y}_{n+1})^T \lambda, \qquad 0 = g(y_{n+1}).$$

We have replaced y_{n+1} with $\tilde{y}_{n+1}$ in the argument of $g'(y)$ in order to save some evaluations of $g'(y)$.

By the middle-value-theorem follows the existence of an x such that

$$\left\| g'(\tilde{y}_{n+1}) - g'(y_{n+1}) \right\| \leq \left\| g''(x) \right\| \left\| \tilde{y}_{n+1} - y_{n+1} \right\|$$

$$\leq C \left\| \tilde{y}_{n+1} - y_{n+1} \right\| = O(h^{p+1})$$

for some $C > 0$.

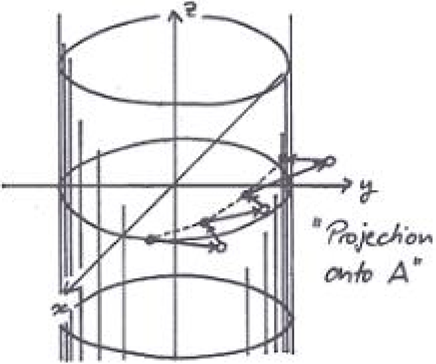
Inserting the first relation ($y_{n+1} = \tilde{y}_{n+1} + g'(\tilde{y}_{n+1})^T\lambda$) into the second ($0 = g(y_{n+1})$) gives a non-linear equation for λ, which can be efficiently solved by **simplified Newton iterations**:

$$\Delta\lambda_i = -(g'(\tilde{y}_{n+1})g'(\tilde{y}_{n+1})^T)^{-1}g(\tilde{y}_{n+1} + g'(\tilde{y}_{n+1})^T\lambda_i),$$

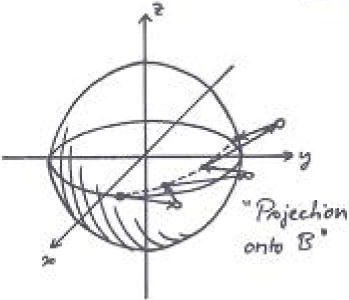
$$\lambda_0 = 0, \quad \lambda_{i+1} = \lambda_i + \Delta\lambda_i.$$

Simplified Newton iteration is Newton iteration with $\tilde{y}_{n+1}$ at some position instead of $\tilde{y}_{n+1} + g'(\tilde{y}_{n+1})^T\lambda$.

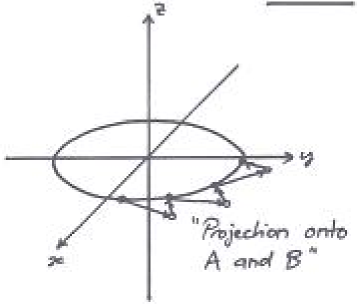
$$g(x,y,z) = x^2 + y^2 - 1 =: \underline{A(x,y,z)}$$



$$g(x,y,z) = x^2+y^2+z^2-1 =: \underline{B(x,y,z)}$$



$$g(x,y,z) = \begin{pmatrix} x^2+y^2-1 \\ x^2+y^2+z^2-1 \end{pmatrix} = \underline{\begin{pmatrix} A(x,y,z) \\ B(x,y,z) \end{pmatrix}}$$



Example (Kepler Problem)

Two first integrals: **Hamiltonian function** $H(q,p)$ and **angular momentum** $L(q,p)$

$$H(q,p) = \frac{1}{2}(p_1^2 + p_2^2) - \frac{1}{\sqrt{q_1^2 + q_2^2}} - \frac{0.005}{2\sqrt{(q_1^2 + q_2^2)^3}},$$

$$L(q,p) = q_1 p_2 - q_2 p_1$$

Initial values: $q_1(0) = 1 - e, \quad q_2(0) = 0,$

$$p_1(0) = 0, \quad p_2(0) = \sqrt{(1+e)/(1-e)}$$

(eccentricity $e = 0.6$)

Remark:

The last term in the Hamiltonian function $-\frac{\mu}{3\sqrt{(q_1^2+q_2^2)^3}}$ is the perturbation term.

- $\mu \neq 0$: **perturbed Kepler problem**, precession of the perihelion
- $\mu = 0$: **Kepler problem**, orbit is an ellipse

Now we discuss the perturbed Kepler problem.

Applied one-step methods:

- explicit Euler: $y_{n+1} = y_n + hf(y_n)$
- symplectic Euler:

$$p_{n+1} = p_n - h\frac{\partial H}{\partial q}(p_{n+1}, q_n), \quad q_{n+1} = q_n + h\frac{\partial H}{\partial p}(p_{n+1}, q_n)$$

Explicit Euler: Projection onto $H(q,p) - H(q_0,p_0)$ has a wrong qualitative behaviour.
Only projection onto both invariants gives the correct motion.
Symplectic Euler: Surprisingly, projection onto $H(q,p) - H(q_0,p_0)$ destroys the correct motion without any projections.
Projection onto both invariants re-establishes the correct behaviour.

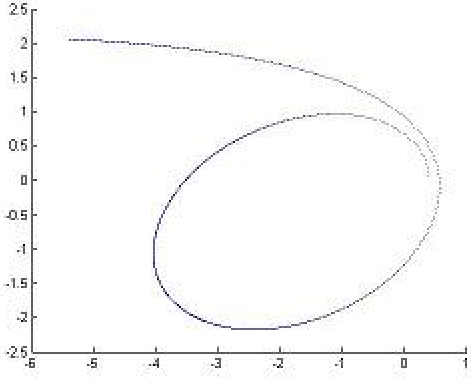


Figure: eE

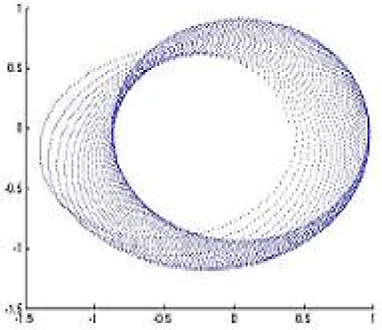


Figure: eEH

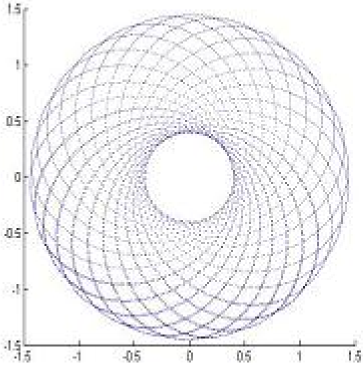


Figure: eEHL

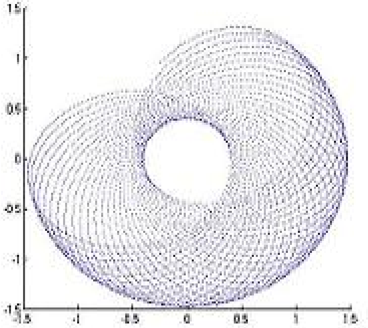


Figure: sE

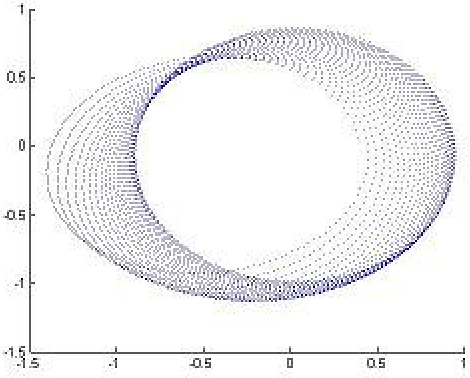


Figure: sEH

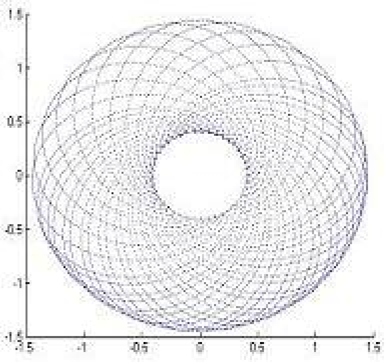


Figure: sEHL

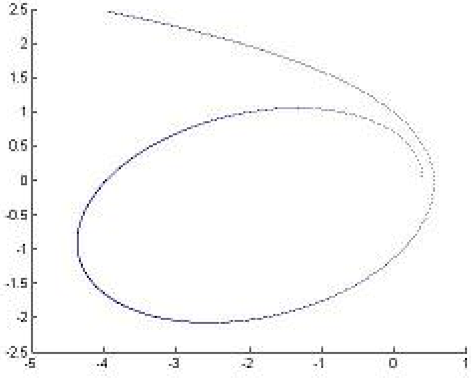


Figure: npeE

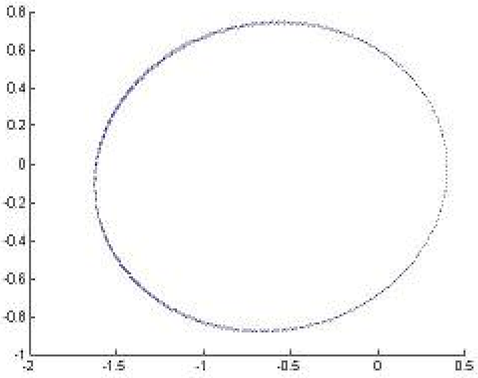


Figure: npsE

Example (Outer Solar System)

Aim: motion of the five planets Jupiter, Saturn, Uranus, Neptune and Pluto relative to the sun. Here q and p are the supervectors composed by the vectors $q_i, p_i \in \mathbb{R}^3, 0 \leq i \leq 5$.

$$H(q,p) = \frac{1}{2}\sum_{i=0}^{5} \frac{1}{m_i} p_i^T p_i - G \sum_{i=1}^{5}\sum_{j=0}^{i-1} \frac{m_i m_j}{\|q_i - q_j\|},$$

$$L(q,p) = \sum_{i=0}^{5} q_i \times p_i,$$

$G \approx 2.96 \cdot 10^{-4}$ is the gravitational constant.

Applying the explicit Euler method with projection onto $H - H_0$ and onto $H - H_0$ and $L - L_0$, we see a slight improvement in the orbits of Jupiter, Saturn and Uranus (compared to the explicit Euler method without projections), but the orbit of Neptune becomes even worse.

This problem contains a structure which cannot be correctly simulated by methods that only preserve the total energy H and the angular momentum L.

In the next two examples we want to compute the **projection step** in concrete problems.

Example (Volume Preservation)

Consider the **matrix differential equation**

$$\dot{Y} = A(Y)Y,$$

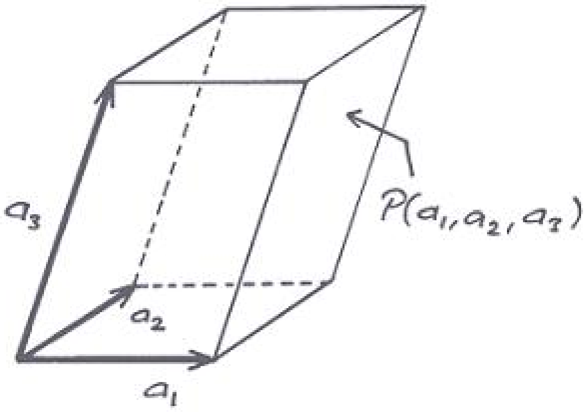
where $trace(A(Y)) = 0$ for all Y.

From last time we know the following Lemma:

Lemma

If $trace(A(Y)) = 0$ for all Y, then $g(Y) := det(Y)$ is an invariant of the matrix differential equation.

$$\textit{Moreover} \quad g'(Y)(BY) = trace(B) \cdot det(Y).$$


a_3
$\mathcal{P}(a_1, a_2, a_3)$
a_2
a_1

Let $a_1, \dots, a_n \in \mathbb{R}^n$.

Definition (Parallelepiped)

$$P(a_1, \dots, a_n) := \left\{ x = \sum_{\nu=1}^{n} t_\nu a_\nu \ : \ t_1, \dots, t_n \in [0, 1] \right\}$$

Theorem

$$Vol(P(a_1, \dots, a_n)) = |det(a_1, \dots, a_n)|$$

$\tilde{Y}_{n+1}$: numerical approximation obtained with an arbitrary one-step method

We consider the **Frobenius norm** $\|Y\|_F = \sqrt{\sum_{i,j} |y_{ij}|^2}$ for measuring the distance to the manifold $\{Y : g(Y) = det(Y_0)\}$.

Lagrange function:

$$L(Y_{n+1}) = \left\| Y_{n+1} - \tilde{Y}_{n+1} \right\|_F^2 /2 - g(Y_{n+1})^T \lambda$$

necessary condition:

$$L'(Y_{n+1})(Q) = 0 \quad \forall Q \in \mathbb{R}^{n \times n}$$

Choose $B \in \mathbb{R}^{n \times n}$ s.t. $B\tilde{Y}_{n+1}$ contains only one non-zero element, for example $(B\tilde{Y}_{n+1})_{ij} = 1 \neq 0$.

Define $h(Y_{n+1}) := \left\| Y_{n+1} - \tilde{Y}_{n+1} \right\|_F^2 / 2$

$$L'(Y_{n+1})(B\tilde{Y}_{n+1}) = h'(Y_{n+1})(B\tilde{Y}_{n+1}) - \lambda g'(\tilde{Y}_{n+1})(B\tilde{Y}_{n+1}) = 0$$

- $h'(Y_{n+1})(B\tilde{Y}_{n+1}) = lim_{\epsilon \to 0} \frac{\frac{1}{2}\left\| Y_{n+1} + \epsilon B\tilde{Y}_{n+1} - \tilde{Y}_{n+1} \right\|_F^2 - \frac{1}{2}\left\| Y_{n+1} - \tilde{Y}_{n+1} \right\|_F^2}{\epsilon} = lim_{\epsilon \to 0} \frac{\epsilon((Y_{n+1})_{ij} - (\tilde{Y}_{n+1})_{ij}) + O(\epsilon^2)}{\epsilon} = (Y_{n+1})_{ij} - (\tilde{Y}_{n+1})_{ij}$
- $B = B\tilde{Y}_{n+1} \cdot \tilde{Y}_{n+1}^{-1}$ is a matrix with non-zero elements only in row i and this row is the row j of $\tilde{Y}_{n+1}^{-1}$,
  $\Rightarrow trace(B) = (\tilde{Y}_{n+1}^{-1})_{ji}$,
  $\Rightarrow g'(\tilde{Y}_{n+1})(B\tilde{Y}_{n+1}) = (\tilde{Y}_{n+1}^{-1})_{ji} \cdot det(\tilde{Y}_{n+1})$

It follows $(Y_{n+1})_{ij} - (\tilde{Y}_{n+1})_{ij} - \lambda(\tilde{Y}_{n+1}^{-T})_{ij} \cdot det(\tilde{Y}_{n+1}) = 0$

and therefore $Y_{n+1} - \tilde{Y}_{n+1} - \lambda \tilde{Y}_{n+1}^{-T} \cdot det(\tilde{Y}_{n+1}) = 0$.

So the **projection step** yields $Y_{n+1} = \tilde{Y}_{n+1} + \mu \tilde{Y}_{n+1}^{-T}$ with $\mu = \lambda det(\tilde{Y}_{n+1})$.

Since one has to solve $g(Y_{n+1}) = g(Y_n)$, this leads to the nonlinear equation $det(Y_n) = det(\tilde{Y}_{n+1} + \mu \tilde{Y}_{n+1}^{-T})$ for μ, for which we apply the **simplified Newton iteration**.

True Newton iteration is $\mu_{i+1} = \mu_i - (f'(\mu_i))^{-1} f(\mu_i)$, where $f(\mu) := g(\tilde{Y}_{n+1} + \mu \tilde{Y}_{n+1}^{-T}) - g(Y_n) = 0$.

$$f'(\mu) = lim_{\epsilon \to 0} \frac{det(\tilde{Y}_{n+1} + (\mu+\epsilon)\tilde{Y}_{n+1}^{-T}) - det(\tilde{Y}_{n+1} + \mu \tilde{Y}_{n+1}^{-T})}{\epsilon} =$$
$$lim_{\epsilon \to 0} \frac{det((\tilde{Y}_{n+1} + \mu \tilde{Y}_{n+1}^{-T})(I + \epsilon(\tilde{Y}_{n+1} + \mu \tilde{Y}_{n+1}^{-T})^{-1} \tilde{Y}_{n+1}^{-T})) - det(\tilde{Y}_{n+1} + \mu \tilde{Y}_{n+1}^{-T})}{\epsilon} =$$
$$lim_{\epsilon \to 0} \frac{det(\tilde{Y}_{n+1} + \mu \tilde{Y}_{n+1}^{-T})(det(I + \epsilon(\tilde{Y}_{n+1} + \mu \tilde{Y}_{n+1}^{-T})^{-1} \tilde{Y}_{n+1}^{-T}) - 1)}{\epsilon} =$$
$$lim_{\epsilon \to 0} \frac{det(\tilde{Y}_{n+1} + \mu \tilde{Y}_{n+1}^{-T})(\epsilon trace((\tilde{Y}_{n+1} + \mu \tilde{Y}_{n+1}^{-T})^{-1} \tilde{Y}_{n+1}^{-T}) + O(\epsilon^2))}{\epsilon} =$$
$$det(\tilde{Y}_{n+1} + \mu \tilde{Y}_{n+1}^{-T}) trace((\tilde{Y}_{n+1} + \mu \tilde{Y}_{n+1}^{-T})^{-1} \tilde{Y}_{n+1}^{-T})$$

So **true Newton iteration** is

$$\Delta\mu_i = \frac{g(Y_n) - g(\tilde{Y}_{n+1} + \mu_i \tilde{Y}_{n+1}^{-T})}{det(\tilde{Y}_{n+1} + \mu_i \tilde{Y}_{n+1}^{-T}) trace((\tilde{Y}_{n+1} + \mu_i \tilde{Y}_{n+1}^{-T})^{-1} \tilde{Y}_{n+1}^{-T})}.$$

Now we take a **simplified version**:

$$\Delta\mu_i = \frac{g(Y_n) - g(\tilde{Y}_{n+1} + \mu_i \tilde{Y}_{n+1}^{-T})}{det(\tilde{Y}_{n+1} + \mu_i \tilde{Y}_{n+1}^{-T}) trace(\tilde{Y}_{n+1}^{-1} \tilde{Y}_{n+1}^{-T})}.$$

We get: $\Delta\mu_i$

$$= \frac{g(Y_n)}{det(\tilde{Y}_{n+1} + \mu_i \tilde{Y}_{n+1}^{-T}) trace((\tilde{Y}_{n+1}^{T} \tilde{Y}_{n+1})^{-1})} - \frac{1}{trace((\tilde{Y}_{n+1}^{T} \tilde{Y}_{n+1})^{-1})},$$

$$\mu_{i+1} = \mu_i + \Delta\mu_i.$$

Example (Orthogonal Matrices)

$\dot{Y} = F(Y)$, where the solution $Y(t)$ is known to be an **orthogonal matrix**, or, more generally, an $n \times k$ matrix $(n \geq k)$ satisfying $Y^T Y = I$ (**Stiefel manifold**).

The **projection step** requires the solution of the problem
$\left\| Y - \tilde{Y} \right\|_F \rightarrow min$ subject to $Y^T Y = I$.

The **projection** can be computed as follows: If $\tilde{Y}$ has the **singular value decomposition** $\tilde{Y} = U^T \Sigma V$, where U^T and V are $n \times k$ and $k \times k$ matrices with orthonormal columns,
$\Sigma = diag(\sigma_1, ..., \sigma_k)$, and the singular values $\sigma_1 \geq ... \geq \sigma_k$ are all close to 1. Then the solution is given by $Y = U^T V$.

We prove the statement for n=k (orthogonal matrices).

Assume $\tilde{Y} = U^T \Sigma V$.

Since $\left\| U^T S V \right\|_F = \|S\|_F$ holds for all orthogonal matrices U and V, it is sufficient to show the case $\|\Sigma - I\|_F = min$ in order to prove $\left\| \tilde{Y} - Y \right\|_F = \left\| U^T \Sigma V - U^T V \right\|_F = min$.

Since $\sigma_i > 0$ close to 1 :
$min_{A \in O(n)} \|\Sigma - A\|_F^2 \quad = \quad min_{A \in O(n), A = diag(\pm 1, \ldots, \pm 1)} \|\Sigma - A\|_F^2 \quad = \|\Sigma - I\|_F^2 \quad = \quad \sum_{i=1}^{n} (\sigma_i - 1)^2.$