UNIVERSITÉ DE GENÈVE

Multiple shooting for Stiefel geodesics

Marco Sutti and Bart Vandereycken

Section of Mathematics, University of Geneva

Overview

- Several applications in optimization, image and signal processing deal with data belonging to the **Stiefel manifold**

$$\mathrm{St}(n,p) = \{X \in \mathbb{R}^{n\times p} : \ X^\top X = I_p\}.$$

- Some applications require evaluating the geodesic distance between two arbitrary points on $\mathrm{St}(n,p)$. No closed-form solution is known for $\mathrm{St}(n,p)$.
- A new computational framework for computing the geodesic distance is proposed, based on the multiple shooting method and the leapfrog algorithm by L. Noakes.
- **Two example applications**:
  - Karcher mean on the space of probability density functions (PDFs);
  - Interpolation of data belonging to $\mathrm{St}(n,p)$ for parametric model reduction.

Geodesics on $\mathrm{St}(n,p)$

- **Geodesic**: generalization of straight lines to manifolds.
- When the tangent space $T_X\mathrm{St}(n,p)$ is endowed with the canonical metric

$$g_c(\Delta,\Delta) = \mathrm{tr}(\Delta^\top(I - \tfrac{1}{2}XX^\top)\Delta), \quad \Delta \in T_X\mathrm{St}(n,p),$$

one can get the following ODE for the geodesic $Z \equiv Z(t)$ [1, eq. (2.41)]:

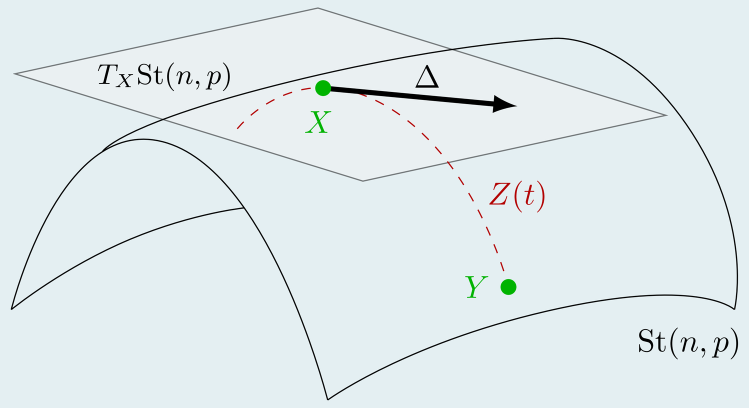
$$\ddot{Z} + \dot{Z}\dot{Z}^\top Z + Z((Z^\top\dot{Z})^2 + \dot{Z}^\top\dot{Z}) = 0.$$

- Closed-form solution for a geodesic $Z(t)$ that realizes a tangent vector Δ with base point X (Ross Lippert [1, eq. (2.42)]):

$$Z(t) = [X \ \ X_\perp] \exp\left(\begin{bmatrix} X^\top\Delta & -(X_\perp^\top\Delta)^\top \\ X_\perp^\top\Delta & O\end{bmatrix} t\right)\begin{bmatrix} I_p \\ O\end{bmatrix}.$$

Riemannian logarithm on $\mathrm{St}(n,p)$

- Given $X, Y \in \mathrm{St}(n,p)$, the **geodesic distance** $d(X,Y)$ is the length of $\Delta_* \equiv \dot{Z}(0) \in T_X\mathrm{St}(n,p)$ s.t. the Riemannian exponential mapping $\mathrm{Exp}_X(\Delta_*) = Y$.
- Equivalent to: Find the **Riemannian logarithm** of Y with base point X, i.e., $\mathrm{Log}_X(Y) = \Delta_*$.



Problem statement: Find $\Delta_* \equiv \dot{Z}(0) \in T_X\mathrm{St}(n,p)$ that satisfies the BVP

$$\ddot{Z} = -\dot{Z}\dot{Z}^\top Z - Z((Z^\top\dot{Z})^2 + \dot{Z}^\top\dot{Z}), \quad \text{with BCs} \begin{cases} Z(0) = X, \\ Z(1) = Y.\end{cases}$$

- **No closed-form solution to this problem is known for $\mathrm{St}(n,p)$!**

Single shooting method

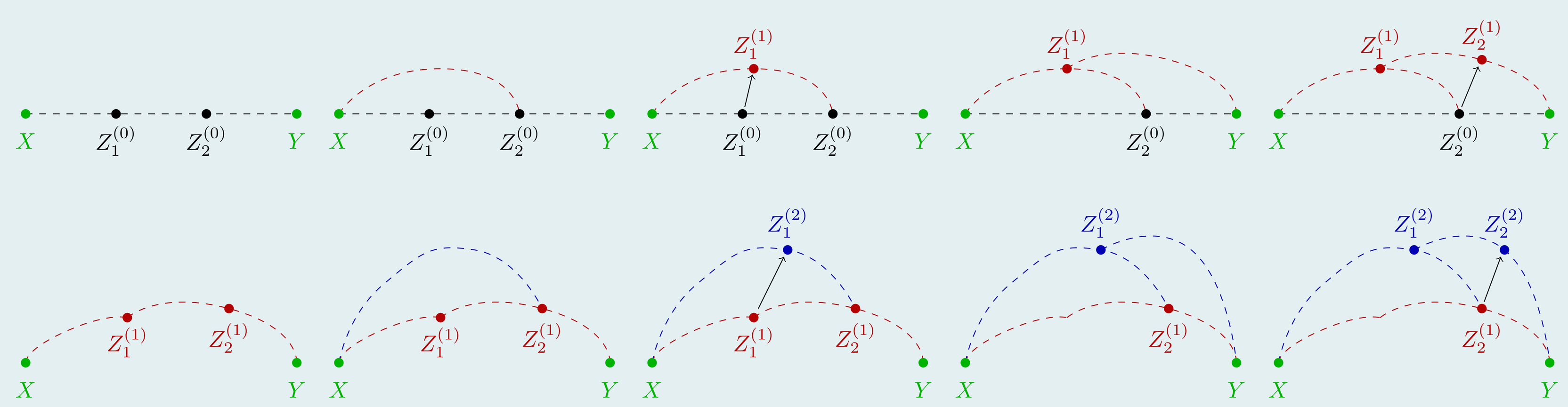
- Define $F(\Delta) = Z_{(t=1,\Delta)} - Y$. Find Δ_* s.t. $F(\Delta_*) = 0$ with **Newton's method**.
- All information is contained in a smaller problem on $\mathrm{St}(2p,p) \longrightarrow$ complexity reduces from $O(n^3)$ to $O(p^3)$ [1].
- A closed-form expression for the Fréchet derivative of the matrix exponential $K^A_{\exp(A)}$ [2, eq. (10.17b)] allows for **explicit expressions of the Jacobians**

$$K^A_{\exp(A)} = (\exp(A^\top/2) \otimes \exp(A/2))\,\mathrm{sinch}\left(\tfrac{1}{2}[A^\top \oplus (-A)]\right).$$

- Fast convergence, but a very good initial guess $\Delta^{(0)}$ is needed.

Leapfrog algorithm (by L. Noakes [3])

- Based on subdivision, s.t. single shooting works well on each subinterval.
- Illustration of two iterations of the procedure, for m points:



- Global convergence to Δ_*, but very slow. Deteriorates when $m \to \infty$.

Multiple shooting method

- Enforce continuity conditions of Z and $\dot{Z}$ at the interfaces between subintervals.
- Fast convergence to Δ_*.
- $\Sigma_1^{(k)}$: point on $\mathrm{St}(n,p)$ relative to the k-th subinterval.
- $\Sigma_2^{(k)}$: tangent vector to $\mathrm{St}(n,p)$ at $\Sigma_1^{(k)}$.

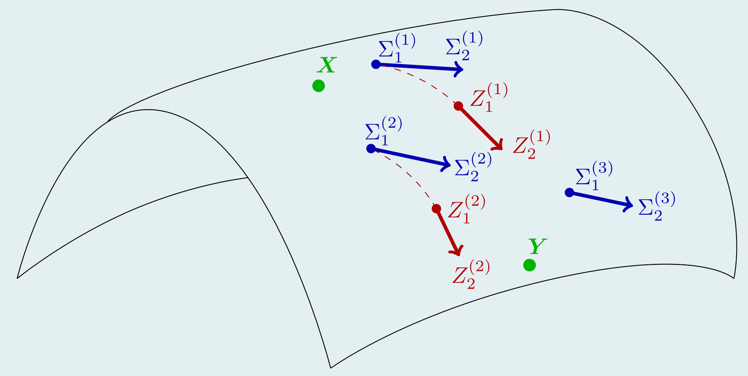


Figure: Multiple shooting on $\mathrm{St}(n,p)$.

System of nonlinear equations:

For each subinterval k, we have an **explicit expression** for the Jacobian $G^{(k)}$.

$$F(\Sigma) := \begin{bmatrix} Z_1^{(1)} - \Sigma_1^{(2)} \\ Z_2^{(1)} - \Sigma_2^{(2)} \\ Z_1^{(2)} - \Sigma_1^{(3)} \\ Z_2^{(2)} - \Sigma_2^{(3)} \\ \vdots \\ r_1 := \Sigma_1^{(1)} - Y_0 \\ r_2 := \Sigma_1^{(m)} - Y_1 \end{bmatrix} = 0, \ \xrightarrow{\text{linearize}} F(\Sigma) + \underbrace{\begin{bmatrix} G^{(1)} & -I & O & & O \\ O & G^{(2)} & -I & \ddots & \\ & \ddots & \ddots & \ddots & O \\ O & & \ddots & G^{(m-1)} & -I \\ C & O & & O & D \end{bmatrix}}_{=:DF(\Sigma)} \delta\Sigma = 0.$$

Our Stiefel Log algorithm: shooting and leapfrog

- To compute the Riemannian logarithm on $\mathrm{St}(n,p)$, single shooting, leapfrog and multiple shooting are combined as illustrated by the flowchart below.

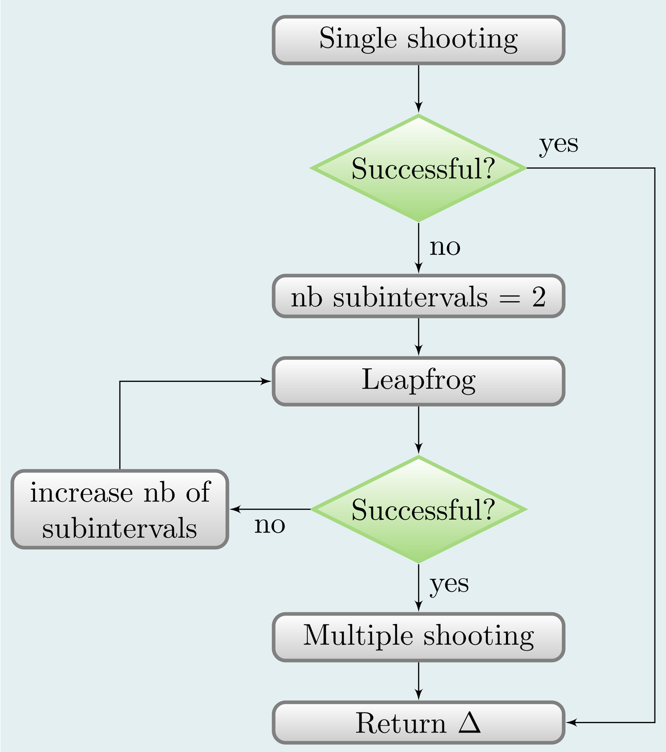
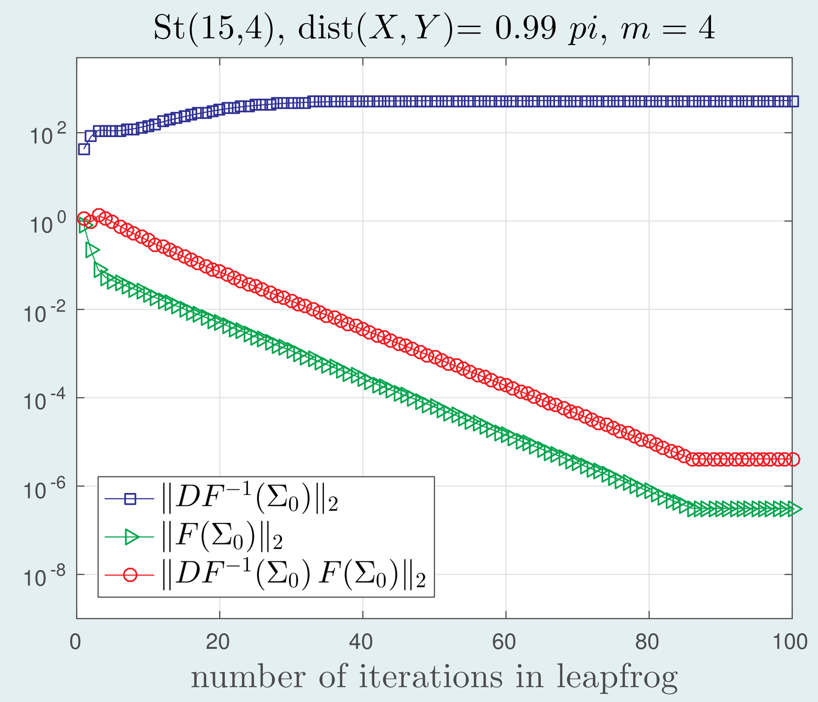


Figure: Flowchart of the Stiefel Log algorithm.

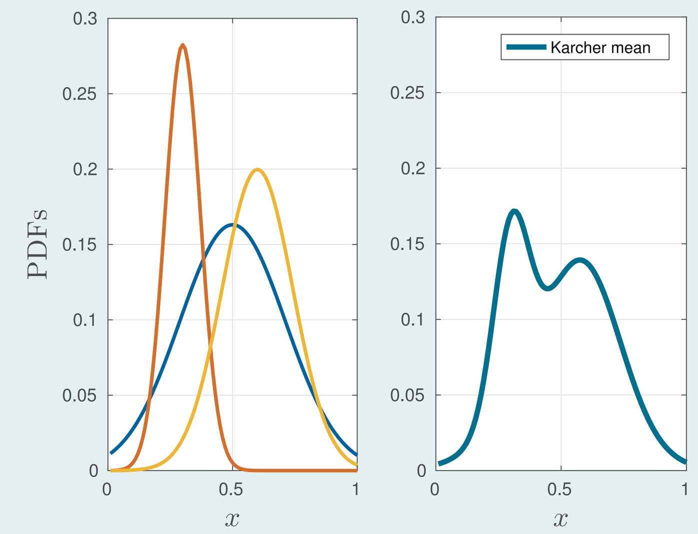
- We observe that $F(\Sigma_0) \to 0$ as the number of iterations in the leapfrog algorithm increases.
- Leapfrog is used to initialize multiple shooting, to enforce the Newton-Kantorovich condition $\|DF(\Sigma_0)^{-1}F(\Sigma_0)\| \leq \alpha$.



- Second N.-K. condition (work in progress): $\|DF(\Sigma_0)^{-1}(DF(\xi) - DF(\zeta))\| \leq \bar{\omega}\|\xi - \zeta\|$.

Karcher mean of univariate probability density functions

- **Karcher mean**: one possible notion of mean on a Riemannian manifold $\mathcal{M}$, defined by the optimization problem $\mu = \arg\min_{p\in\mathcal{M}} \frac{1}{2N}\sum_{i=1}^N d(p,q_i)^2$, where $d(p,q_i)$ is the distance between two points on $\mathcal{M}$.
- $\mathcal{S}^n = \{x \in \mathbb{R}^{n+1} : \|x\| = 1\}$ can be used to approximate $\mathcal{S}^\infty$, which represents the **space of univariate PDFs** on the unit interval $[0,1]$, i.e., $\mathcal{P} = \{g : [0,1] \to \mathbb{R}_{\geq 0} : \int_0^1 g(x)\,\mathrm{d}x = 1\}$.
- **Example**: Karcher mean of 3 PDFs, sampled at 100 points, which makes them belonging to $\mathrm{St}(100,1)$.



Model reduction with POD and interpolation on $\mathrm{St}(n,r)$

- Model reduction for dynamical systems parametrized with $\mathbf{p} = [p_1,\dots,p_d]^\top$:

$$\begin{cases} \dot{\mathbf{x}}(t;\mathbf{p}) = \mathbf{A}(\mathbf{p})\,\mathbf{x}(t;\mathbf{p}) + \mathbf{B}(\mathbf{p})\,\mathbf{u}(t), \\ \mathbf{y}(t;\mathbf{p}) = \mathbf{C}(\mathbf{p})\,\mathbf{x}(t;\mathbf{p}), \end{cases} \xrightarrow{\text{reduction}} \begin{cases} \dot{\mathbf{x}}_r(t;\mathbf{p}) = \mathbf{A}_r(\mathbf{p})\,\mathbf{x}_r(t;\mathbf{p}) + \mathbf{B}_r(\mathbf{p})\,\mathbf{u}(t), \\ \mathbf{y}_r(t;\mathbf{p}) = \mathbf{C}_r(\mathbf{p})\,\mathbf{x}_r(t;\mathbf{p}), \end{cases}$$

$\mathbf{x}(t;\mathbf{p}) \in \mathbb{R}^n$, $\mathbf{u}(t) \in \mathbb{R}^m$, $\mathbf{y}(t) \in \mathbb{R}^q$, $\mathbf{A}(\mathbf{p}) \in \mathbb{R}^{n\times n}$, $\mathbf{B}(\mathbf{p}) \in \mathbb{R}^{n\times m}$, $\mathbf{C}(\mathbf{p}) \in \mathbb{R}^{q\times n}$.

$\mathbf{x}_r = \mathbf{V}^\top\mathbf{x}$, $\mathbf{A}_r = \mathbf{V}^\top\mathbf{A}\mathbf{V}$, $\mathbf{B}_r = \mathbf{V}^\top\mathbf{B}$, $\mathbf{C}_r = \mathbf{C}\mathbf{V}$, $\mathbf{V} \equiv \mathbf{V}(\mathbf{p}) \in \mathrm{St}(n,r)$.

- For each parameter in a set of parameter values $\{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_K\}$, use proper orthogonal decomposition (POD) to derive a reduced-order basis $\mathbf{V}_i \in \mathrm{St}(n,r)$.

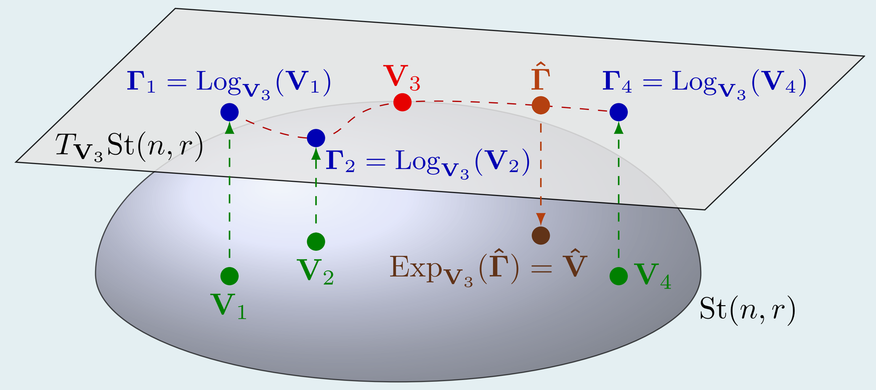
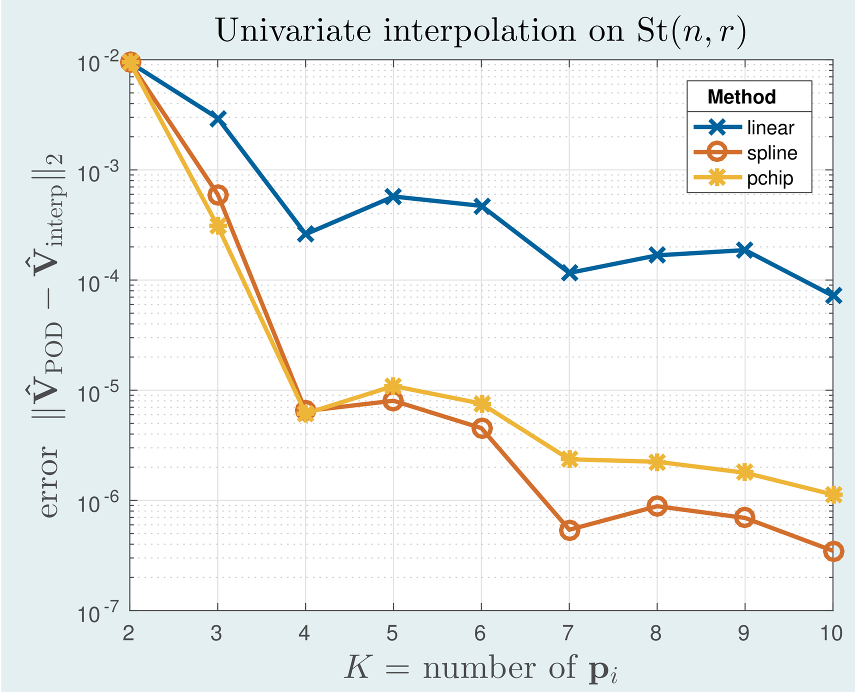
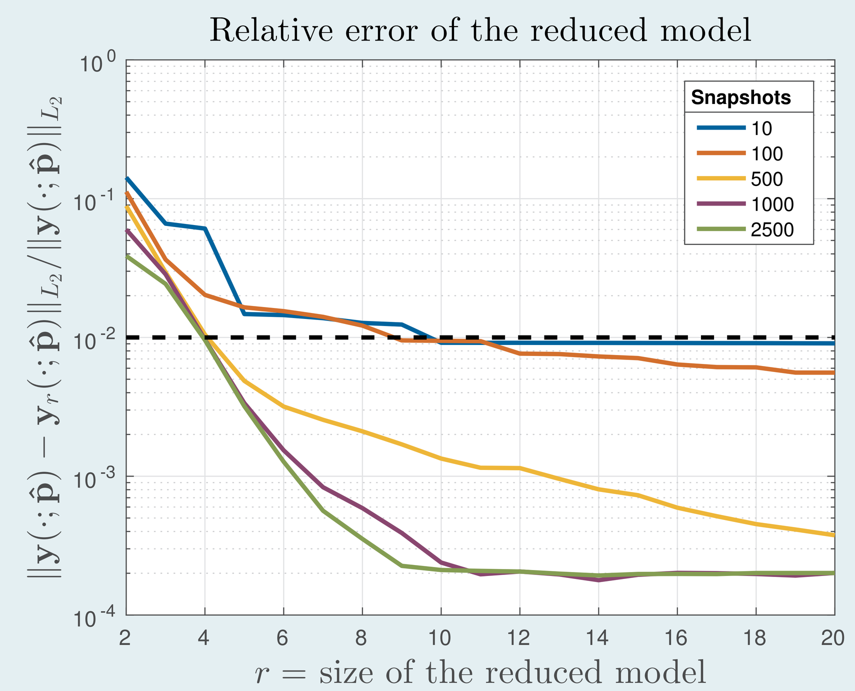


Figure: Interpolation on $\mathrm{St}(n,r)$.

- This yields a set of local basis matrices $\{\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_K\}$.
- Given a new parameter value $\hat{\mathbf{p}}$, a basis $\hat{\mathbf{V}}$ can be obtained by **interpolating the local basis matrices on a tangent space to $\mathrm{St}(n,r)$**.





- **Application**: transient heat equation on a square domain, with 4 disjoint discs.
- FEM discretization with $n = 1169$. Simulation for $t \in [0,500]$, with $\Delta t = 0.1$.
- 500 snapshot POD over 5000 timeframes, with a reduced model of size $r = 4$.
- Relative error between $\mathbf{y}(\cdot;\hat{\mathbf{p}})$ and $\mathbf{y}_r(\cdot;\hat{\mathbf{p}})$ is about 1%.

Essential references

[1] Alan Edelman, Toms A. Arias, and Steven T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications*, 20(2):303–353, 1998.

[2] Nicholas J. Higham. *Functions of Matrices: Theory and Computation*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2008.

[3] Lyle Noakes. A global algorithm for geodesics. *Journal of the Australian Mathematical Society. Series A. Pure Mathematics and Statistics*, 65(1):37–50, 008 1998.