MANCHESTER 1824

The University of Manchester

M\cr NA
Manchester Numerical Analysis

The Structured Condition Number of a Differentiable Map Between Matrix Manifolds

Françoise Tisseur
School of Mathematics
The University of Manchester

Joint work with
Bahar Arslan (Bursa Technical University)
Vanni Noferini (University of Essex)

INdAM Meeting, Cortona, September 2017

Motivating Example

Let $a > 0$ and consider the matrix logarithm of X,

$$X = \begin{bmatrix} e^a & 0 \\ 0 & e^{-a} \end{bmatrix}, \qquad \log X = \begin{bmatrix} a & 0 \\ 0 & -a \end{bmatrix}.$$

X symplectic $\Rightarrow$ $\log(X)$ is Hamiltonian.

Motivating Example

Let $a > 0$ and consider the matrix logarithm of X,

$$X = \begin{bmatrix} e^a & 0 \\ 0 & e^{-a} \end{bmatrix}, \qquad \log X = \begin{bmatrix} a & 0 \\ 0 & -a \end{bmatrix}.$$

X symplectic $\Rightarrow$ $\log(X)$ is Hamiltonian.

Apply tiny perturbation to X (norm = $\epsilon\|X\|$),

$$\widetilde{X} = \begin{bmatrix} e^a & 0 \\ 0 & e^{-a} + \epsilon e^a \end{bmatrix} \Rightarrow \log \widetilde{X} = \begin{bmatrix} a & 0 \\ 0 & -a \end{bmatrix} + \epsilon \begin{bmatrix} 0 & 0 \\ 0 & e^{2a} \end{bmatrix} + o(\epsilon).$$

Worst case scenario, so the **relative conditioning** is

$$\frac{e^{2a}}{a}.$$

Motivating Example (Cont.)

Now consider a pertubation preserving the symplectic structure,

$$X = \begin{bmatrix} e^a & \epsilon e^a \\ 0 & e^{-a} \end{bmatrix} \Rightarrow \ln X = \begin{bmatrix} a & 0 \\ 0 & -a \end{bmatrix} + \epsilon \begin{bmatrix} 0 & \frac{ae^a}{\sinh(a)} \\ 0 & 0 \end{bmatrix} + o(\epsilon)$$

This is the worse we can do but the effect is much milder!

The **relative structured conditioning** is

$$\frac{e^a}{\sinh(a)} < \frac{e^{2a}}{a}$$

Structure conditioning $\ll$ unstructured conditioning for average or large values of a.

Condition Number of $f(X)$

Let $f : \mathbb{F}^{n\times n} \to \mathbb{F}^{n\times n}$ be $\mathbb{F}$-Fréchet differentiable, $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$.

Define absolute **condition number of $f(X)$**:

$$\mathrm{cond}(f, X) = \lim_{\epsilon\to 0} \sup_{\substack{\|Y-X\|\le\epsilon \\ Y\in\mathbb{F}^{n\times n}}} \frac{\|f(Y) - f(X)\|}{\epsilon}.$$

Condition Number of $f(X)$

Let $f : \mathbb{F}^{n\times n} \to \mathbb{F}^{n\times n}$ be $\mathbb{F}$-Fréchet differentiable, $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$.

Define absolute **condition number of $f(X)$**:

$$\mathrm{cond}(f, X) = \lim_{\epsilon\to 0} \sup_{\substack{\|Y-X\|\le\epsilon \\ Y\in\mathbb{F}^{n\times n}}} \frac{\|f(Y) - f(X)\|}{\epsilon}.$$

The Fréchet derivative of f at X is a $\mathbb{F}$-linear map $L_f(X, \cdot)$ so

$$\mathrm{vec}(L_f(X, E)) = K_f(X)\mathrm{vec}(E),$$

where $K_f(X) \in \mathbb{F}^{n^2\times n^2}$ is **Kronecker form of the Fréchet derivative**. Then, for the Frobenius norm,

$$\mathrm{cond}(f, X) = \max_{\substack{E\in\mathbb{F}^{n\times n} \\ E\neq 0}} \frac{\|L_f(X, E)\|_F}{\|E\|_F} = \|K_f(X)\|_2 .$$

Restricting the Perturbations

Suppose $f : \mathcal{M} \to \mathcal{N}$ is differentiable with $\mathcal{M}, \mathcal{N} \subseteq \mathbb{F}^{n\times n}$ **smooth matrix manifolds**. Let $X \in \mathcal{M}$ and enforce the perturbations $Y - X$ to be s.t. $Y \in \mathcal{M}$.

Define **structured condition number of f at X**:

$$\mathrm{cond}_{\mathsf{struc}}(f, X) = \lim_{\epsilon\to 0} \sup_{\substack{\|Y-X\|\le\epsilon \\ Y\in\mathcal{M}}} \frac{\|f(Y) - f(X)\|}{\epsilon}$$

$$\Rightarrow \|f(Y) - f(X)\| \le \mathrm{cond}_{\mathsf{struc}}(f, X)\|Y - X\| + o(\|Y - X\|).$$

Also, by the definition of supremum, we have

$$\mathrm{cond}_{\mathsf{struc}}(f, X) \le \mathrm{cond}(f, X).$$

Square Matrix Manifolds

Consider three classes:

1. real submanifolds of the n^2-dimensional real vector space $\mathbb{R}^{n\times n}$ (e.g., real symplectic matrices),
2. complex submanifolds of the n^2-dimensional complex vector space $\mathbb{C}^{n\times n}$ (e.g., complex orthogonal matrices),
3. real submanifolds of the $2n^2$-dimensional *real* vector space $\mathbb{C}^{n\times n}$ (e.g., Hermitian, unitary matrices).

Need to distinguish between

- **ambient field** $\mathbb{F}$: field in which the entries of X lie,
- **base field** $\mathbb{K}$: field the ambient vector space is built on.

$$X = X^* = \begin{bmatrix} a & b + \mathrm{i}c \\ b - \mathrm{i}c & d \end{bmatrix},\ a, b, c, d \in \mathbb{R} \text{ so } \mathbb{F} = \mathbb{C},\ \mathbb{K} = \mathbb{R}.$$

Theoretical Tools: Differential Geometry

Let $\mathcal{M}, \mathcal{N} \subseteq \mathbb{F}^{n\times n}$ with base field $\mathbb{K}$ and let $f : \mathcal{M} \to \mathcal{N}$ be $\mathbb{K}$-differentiable.

(*The differential plays the role of the derivative for maps f between manifolds: best (local)* $\mathbb{K}$*-linear approx. to f.*)

$$df_X : T_X\mathcal{M} \to T_{f(X)}\mathcal{N}, \quad df_X(\gamma'(0)) = (f \circ \gamma)'(0),$$

for a smooth curve $\gamma : \mathbb{K} \to \mathcal{U} \subset \mathcal{M}$ such that $\gamma(0) = X$,
$T_X\mathcal{M}$: **tangent space** of $\mathcal{M}$ at X.

Theorem (Arslan, Noferini, T.'17)

For the ***structured condition number of*** f ***at*** $X \in \mathcal{M}$ *we have*

$$\mathrm{cond}_{\text{struc}}(f, X) = \|df_X\|.$$

Structured Condition Number (Cont.)

Uniqueness of differential and Fréchet derivative imply $\Rightarrow$

$$df_X(E) = L_f(X, E), \quad E \in T_X\mathcal{M}.$$

Structured Condition Number (Cont.)

Uniqueness of differential and Fréchet derivative imply $\Rightarrow$

$$df_X(E) = L_f(X, E), \quad E \in T_X\mathcal{M}.$$

$T_X\mathcal{M}$ is a **$\mathbb{K}$-linear** subspace of $\mathbb{F}^{n\times n}$ so $\text{vec}(E) = By$,
$B \in \mathbb{F}^{n^2\times p}$ (basis for $T_X\mathcal{M}$), $y \in \mathbb{K}^p$ (vector of parameters),
$p = \dim_{\mathbb{K}} T_X\mathcal{M}$.

Structured Condition Number (Cont.)

Uniqueness of differential and Fréchet derivative imply $\Rightarrow$

$$df_X(E) = L_f(X, E), \quad E \in T_X\mathcal{M}.$$

$T_X\mathcal{M}$ is a $\mathbb{K}$**-linear** subspace of $\mathbb{F}^{n\times n}$ so $\text{vec}(E) = By$,
$B \in \mathbb{F}^{n^2\times p}$ (basis for $T_X\mathcal{M}$), $y \in \mathbb{K}^p$ (vector of parameters),
$p = \dim_{\mathbb{K}} T_X\mathcal{M}$. Now using

$$\text{vec}(df_X(E)) = \text{vec}(L_f(X, E)) = K_f(X)\text{vec}(E) = K_f(X)By,$$

we have

$$\text{cond}_{\text{struc}}(f, X) = \max_{\substack{E\in T_X\mathcal{M}\\ E\neq 0}} \frac{\|df_X(E)\|_F}{\|E\|_F} = \max_{\substack{y\in\mathbb{K}^p\\ y\neq 0}} \frac{\|K_f(X)By\|_2}{\|By\|_2}.$$

If $B^*B = I_p$ then $\text{cond}_{\text{struc}}(f, X) = \|K_f(X)B\|_2$.

From Theory to Computation

The theory is simple and elegant, but computing the (un)structured condition number is an issue, since the cost of a naive algorithm is

- $O(n^5)$ for $\mathrm{cond}(f, X) = \|K_f(X)\|_2$,
  $K_f(X)e_{i+(j-1)n} = \mathrm{vec}(L_f(X, e_i e_j^T))$, $i = 1\colon n$, $j = 1\colon n$,
- $O(n^6)$ for $\mathrm{cond}_{\mathrm{struc}}(f, X) = \|K_f(X)B\|_2$ when $B^*B = I_p$,
  $K_f(X)Be_i = \mathrm{vec}\big(L_f(X, \mathrm{unvec}(Be_i))\big)$, $i = 1\colon p$.

From Theory to Computation

The theory is simple and elegant, but computing the (un)structured condition number is an issue, since the cost of a naive algorithm is

- $O(n^5)$ for $\mathrm{cond}(f, X) = \|K_f(X)\|_2$,
  $K_f(X)e_{i+(j-1)n} = \mathrm{vec}(L_f(X, e_i e_j^T))$, $i = 1\colon n$, $j = 1\colon n$,
- $O(n^6)$ for $\mathrm{cond}_{\mathrm{struc}}(f, X) = \|K_f(X)B\|_2$ when $B^*B = I_p$,
  $K_f(X)Be_i = \mathrm{vec}\big(L_f(X, \mathrm{unvec}(Be_i))\big)$, $i = 1\colon p$.

When $B^*B \neq I_p$, can compute the lower/upper bounds

$$\|K_f(X)B\|_2 \|B\|_2^{-1} \leq \mathrm{cond}_{\mathrm{struc}}(f, X) \leq \|K_f(X)B\|_2 \|B^+\|_2.$$

Computational tool: the power method

This alg. applies the power method to $K_f(X)^* B^* B K_f(X)$ to compute a lower bound γ for $\|K_f(X)B\|_2$.

```
Start with z_0 ∈ F^p.
for k = 0: ∞
    vec(E_k) = B z_k
    W_{k+1} = L_f(X, E_k)
    E_{k+1} = L_f^⋆(X, W_{k+1})
    γ_{k+1} = ‖Z_{k+1}‖_F / ‖W_{k+1}‖_F
    z_{k+1} = B^* vec(Y_{k+1})
    if converged, γ = γ_{k+1}, quit, end
end
```

$$L_f^{\star}(X,E) = \begin{cases} L_f(X^T, E), & \mathbb{F} = \mathbb{R} \\ L_{\bar{f}}(X^*, E), & \mathbb{F} = \mathbb{C} \end{cases}$$

Cost: $O(n^2 p)$ per iteration but often the structure of B can be exploited to reduce the cost to $O(np)$. Typically, $p = O(n^2)$.

Structured Matrices

A scalar product $\langle \cdot, \cdot \rangle_M$ is a non degenerate (M nonsingular) bilinear or sesquilinear form on $\mathbb{F}^n$.

$$\langle x, y \rangle_M = \begin{cases} x^T M y & \text{real or complex } \textbf{bilinear} \text{ forms,} \\ x^* M y & \textbf{sesquilinear} \text{ forms.} \end{cases}$$

Any matrix A has a unique adjoint $A^\star$ given by

$$A^\star = \begin{cases} M^{-1} A^T M, & \text{real or complex } \textbf{bilinear} \text{ forms,} \\ M^{-1} A^* M, & \textbf{sesquilinear} \text{ forms.} \end{cases}$$

Structured Matrices

A scalar product $\langle \cdot, \cdot \rangle_M$ is a non degenerate (M nonsingular) bilinear or sesquilinear form on $\mathbb{F}^n$.

$$\langle x, y \rangle_M = \begin{cases} x^T M y & \text{real or complex } \textbf{bilinear} \text{ forms,} \\ x^* M y & \textbf{sesquilinear} \text{ forms.} \end{cases}$$

Any matrix A has a unique adjoint $A^\star$ given by

$$A^\star = \begin{cases} M^{-1} A^T M, & \text{real or complex } \textbf{bilinear} \text{ forms,} \\ M^{-1} A^* M, & \textbf{sesquilinear} \text{ forms.} \end{cases}$$

Associated with $\langle \cdot, \cdot \rangle_M$ are three classes of structured matrices: a Jordan algebra $\mathbb{J}_M$, a Lie algebra $\mathbb{L}_M$, and an automorphism group $\mathbb{G}_M$.

Automorphism Groups

Having fixed the scalar product $\langle \cdot, \cdot \rangle_{\mathrm{M}}$, the associated **automorphism group** is

$$\begin{aligned} \mathbb{G}_{\mathrm{M}} &:= \{G \in \mathbb{F}^{n\times n} : \langle Gx, Gy \rangle_{\mathrm{M}} = \langle x, y \rangle_{\mathrm{M}} \} \\ &= \{G \in \mathbb{F}^{n\times n} : G^{\star} = G^{-1}\}. \end{aligned}$$

Automorphism Groups

Having fixed the scalar product $\langle\cdot,\cdot\rangle_{\mathrm{M}}$, the associated **automorphism group** is

$$\begin{aligned}\mathbb{G}_{\mathrm{M}} :&= \{G \in \mathbb{F}^{n\times n} : \langle Gx, Gy\rangle_{\mathrm{M}} = \langle x, y\rangle_{\mathrm{M}}\} \\ &= \{G \in \mathbb{F}^{n\times n} : G^{\star} = G^{-1}\}.\end{aligned}$$

Theorem (Arslan, Noferini, T.'17)

$\mathbb{G}_{M}$ is a real submanifold of $\mathbb{F}^{n\times n}$. When $\langle\cdot,\cdot\rangle_{M}$ is a complex bilinear form, $\mathbb{G}_{M}$ is also a complex submanifold of $\mathbb{C}^{n\times n}$.

Tangent Spaces and Lie Algebras

Having fixed $\langle\cdot,\cdot\rangle_{\mathrm{M}}$, the associated **Lie algebra** is

$$\begin{aligned}\mathbb{L}_{\mathrm{M}} &:= \{F \in \mathbb{K}^{n\times n} : \langle Fx, y\rangle_{\mathrm{M}} = -\langle x, Fy\rangle_{\mathrm{M}}\} \\ &= \{G \in \mathbb{K}^{n\times n} : G^{\star} = -G\}.\end{aligned}$$

Tangent Spaces and Lie Algebras

Having fixed $\langle\cdot,\cdot\rangle_{\mathrm{M}}$, the associated **Lie algebra** is

$$\mathbb{L}_{\mathrm{M}} := \{F \in \mathbb{K}^{n\times n} : \langle Fx, y\rangle_{\mathrm{M}} = -\langle x, Fy\rangle_{\mathrm{M}}\} \\ = \{G \in \mathbb{K}^{n\times n} : G^{\star} = -G\}.$$

Theorem

Let $X \in \mathbb{G}_M$, then

$$T_X\mathbb{G}_M = \{E \in \mathbb{F}^{n\times n} : E = XF, F \in \mathbb{L}_M.\}.$$

A basis for $T_X\mathbb{G}_M$ is $B = (I_n \otimes XM^{-1})D$, where D has orthonormal columns and does not depend on X.

Tangent Spaces and Lie Algebras

Having fixed $\langle\cdot,\cdot\rangle_{\mathrm{M}}$, the associated **Lie algebra** is

$$\begin{aligned}\mathbb{L}_{\mathrm{M}} &:= \{F \in \mathbb{K}^{n\times n} : \langle Fx, y\rangle_{\mathrm{M}} = -\langle x, Fy\rangle_{\mathrm{M}}\} \\ &= \{G \in \mathbb{K}^{n\times n} : G^{\star} = -G\}.\end{aligned}$$

Theorem

Let $X \in \mathbb{G}_M$, then

$$T_X\mathbb{G}_M = \{E \in \mathbb{F}^{n\times n} : E = XF, F \in \mathbb{L}_M.\}.$$

A basis for $T_X\mathbb{G}_M$ is $B = (I_n \otimes XM^{-1})D$, where D has orthonormal columns and does not depend on X.

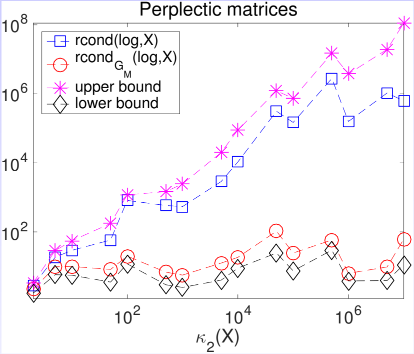
$$\frac{\|K_f(X)B\|_2}{\|M^{-1}\|_2\|X\|_2} \le \mathrm{cond}_{\mathrm{s}}(f, X) \le \|K_f(X)B\|_2\|X\|_2\|M\|_2.$$

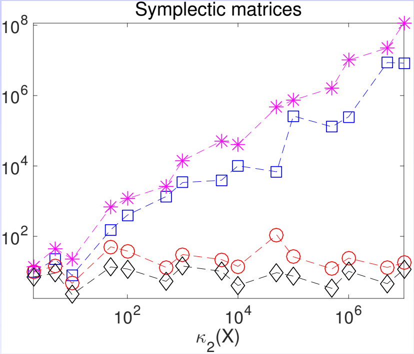
Commonly Used M

M	Structure of M	Name of $\mathbb{G}_M$	Where they appear
I	I_n	orthogonal cplx orthogonal unitary	everywhere
J	$\begin{bmatrix} 0 & I_{n/2} \\ -I_{n/2} & 0 \end{bmatrix}$	symplectic cplx symplectic conjugate sympl.	physics engineering
R	$\begin{bmatrix} & & 1 \\ & ⋰ & \\ 1 & & \end{bmatrix}$	perplectic	maths, chemistry
S_{pq}	$\begin{bmatrix} I_p & 0 \\ 0 & -I_q \end{bmatrix}$ $p+q=n$	pseudo-orthogonal pseudo-unitary	physics

Principal Logarithm

$\log(X) : \mathbb{G}_M \to \mathbb{L}_M$. Generate random $X \in \mathbb{G}_M \subset \mathbb{R}^{10\times 10}$, $M = R, J$ with increasing $\kappa_2(X) = \|X\|_2 \|X^{-1}\|_2$.





$$X^{-1} = RX^T R$$

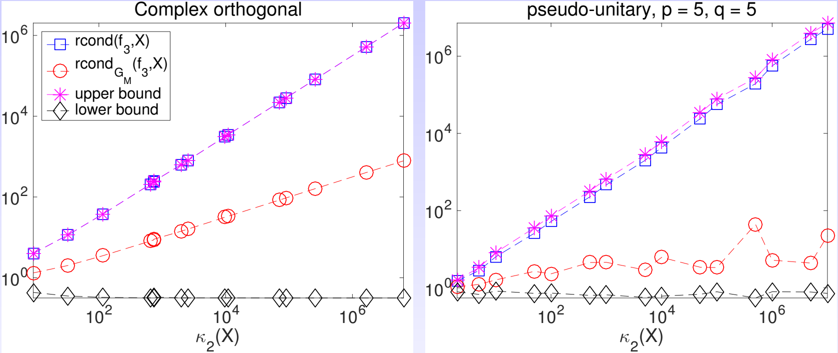
$$R = \begin{bmatrix} & & 1 \\ & \cdot^{\cdot^{\cdot}} & \\ 1 & & \end{bmatrix}$$

$$X^{-1} = -JX^T J$$

$$J = \begin{bmatrix} 0 & I_{n/2} \\ -I_{n/2} & 0 \end{bmatrix}$$

Unitary Polar Factor

$f_3 : \mathbb{G}_M \to \mathbb{G}_M \cap \mathbb{G}_I$, $f_3(X) = U$, where $X = UH$ (polar decomposition). f_3 is only $\mathbb{R}$-differentiable.



$$X^{-1} = X^T, \qquad X^{-1} = \begin{bmatrix} I_p & 0 \\ 0 & -I_q \end{bmatrix} X^* \begin{bmatrix} I_p & 0 \\ 0 & -I_q \end{bmatrix}.$$

For complex orthogonal, must choose base field $\mathbb{K} = \mathbb{R}$.

Principal Square Root $X^{1/2} : \mathbb{G}_M \to \mathbb{G}_M$

Let $A = X^{1/2}$ and $\widehat{A}$ be the computed square root of X.
Relative error bound:

$$\frac{\|\widehat{A} - X^{1/2}\|_F}{\|X^{1/2}\|_F} \lesssim \mathrm{cond}(\mathrm{sqrt}, X) \frac{\|\widehat{A}^2 - X\|_F}{\|X^{1/2}\|_F}. \tag{1}$$

Principal Square Root $X^{1/2} : \mathbb{G}_M \to \mathbb{G}_M$

Let $A = X^{1/2}$ and $\widehat{A}$ be the computed square root of X. Relative error bound:

$$\frac{\|\widehat{A} - X^{1/2}\|_F}{\|X^{1/2}\|_F} \lesssim \mathrm{cond}(\mathrm{sqrt}, X) \frac{\|\widehat{A}^2 - X\|_F}{\|X^{1/2}\|_F}. \qquad (1)$$

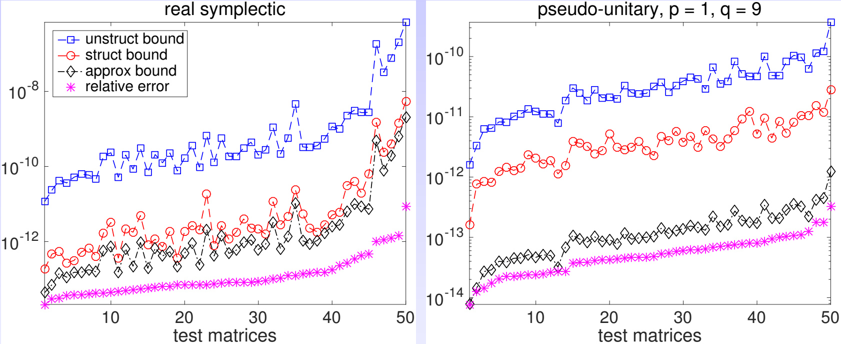
Use structure preserving and cubically converging iteration

$$Y_{k+1} = \frac{1}{3} Y_k \left[I + 8\left(I + 3Z_k Y_k\right)^{-1} \right], \qquad Y_0 = X,$$
$$Z_{k+1} = \frac{1}{3} \left[I + 8\left(I + 3Z_k Y_k\right)^{-1} \right] Z_k, \qquad Z_0 = I,$$

where $Y_k, Z_k \in \mathbb{G}_M$ and $Y_k \to X^{1/2}$ (see Higham, Mackey, Mackey, T., SIMAX'05). Can replace $\mathrm{cond}(\mathrm{sqrt}, X)$ by $\mathrm{cond}_{\mathbb{G}_M}(\mathrm{sqrt}, X)$ in (1).

Relative Error Bounds

Bounds on $\|\widehat{A} - X^{1/2}\|_F / \|X^{1/2}\|_F$, $\kappa_2(X) = 10^5$, $N = 10$.



$$X^{-1} = -JX^TJ \qquad X^{-1} = S_{pq}X^*S_{pq}$$

$$J = \begin{bmatrix} 0 & I_{n/2} \\ -I_{n/2} & 0 \end{bmatrix} \qquad S_{pq} = \begin{bmatrix} I_p & 0 \\ 0 & -I_q \end{bmatrix}.$$

Concluding Remarks

- Framework to define and compute structured condition numbers of maps between matrix manifolds
- Structured condition number expensive to compute but lower/upper bounds are cheap. The lower bound is often a good approximation.
- Gave numerical evidence that often s.c.n. $\ll$ u.c.n. This suggests that using a structure preserving alg to compute $f(X)$ might be advantageous.
- MIMS EPrint available (soon) on my web page.