Low rank approximation for the numerical simulation of high dimensional Riccati equations

Pierre Rouchon
Centre Automatique et Systèmes, Mines-ParisTech

Based on the arXiv preprint
Low rank approximation for the numerical simulation of high dimensional Lindblad and Riccati equations
http://arxiv.org/abs/1207.4580
with Claude Le Bris from CERMICS/INRIA

Colloque National sur l'Assimilation de Données
17, 18 et 19 décembre 2012, Nice.

Outline

Continuous time Kalman filter and large state dimension n

Orthogonal projection on the sub-manifold of rank $r < n$ symmetric matrices

Numerical tests on a toy example: 1D wave equation

Kalman filter: continuous time formulation

Take the linear system of state $x \in \mathbb{R}^n$ and output $y \in \mathbb{R}^m$

$$dx = Ax\,dt + G\,d\eta, \quad dy = C\,dx + H\,d\mu$$

where η and μ are independent Wierner processes. Assume that H is invertible.
The expectation value $\widehat{x}$ of $x(t)$ knowing $y([0,t])$, $\widehat{x}_0$ and $P_0 = \mathbb{E}\left(|x-\widehat{x}\rangle\langle x-\widehat{x}|\right)_{t=0}$ is given by the asymptotic observer (Kalman filter)

$$\frac{d}{dt}\widehat{x} = A\widehat{x} + L(C\widehat{x} - y(t)), \quad \widehat{x}(0) = \widehat{x}_0$$

where the gain $L(t) = -P(t)C^T(HH^T)^{-1}$ relies on the covariance matrix P, solution of the Riccati differential equation

$$\frac{d}{dt}P = AP + PA^T + GG^T - PC^T(HH^T)^{-1}CP, \quad P(0) = P_0.$$

Extended Kalman Filter (EKF) for $\frac{d}{dt}x = f(x,t)$, $y = h(x,t)$:

$$\frac{d}{dt}\widehat{x} = f(\widehat{x},t) + L(h(\widehat{x},t) - y(t))$$

with $A(t) = \frac{\partial f}{\partial x}(\widehat{x}(t),t)$, $C(t) = \frac{\partial h}{\partial x}(\widehat{x}(t),t)$.

Filtering for high dimensional systems

For state space dimension n very large, the manipulation and storage of the covariance $n \times n$ matrix P become problematic. Several approximations have been developed:

- Reduced order filters based on a reduction in the order n of the system model.
- Ensemble Kalman filters
- Particle filters
- SEEK filters
- ...

Key issue: the sub-space spanned by the eigen-vectors corresponding to the largest eigenvalues of P; how this sub-space evolves (rotates) versus time ?

Contribution(if not already done elsewhere): use the orthogonal projection of the Riccati differential equation onto the tangent space to the sub-manifold of rank r symmetric matrices.

Low rank solution of the Riccati differential equation[1]

$$dx = Ax\,dt + G\,d\eta, \quad dy = C\,dx + H\,d\mu:$$

$$\frac{d}{dt}P = AP + PA^T + GG^T - PC^T(HH^T)^{-1}CP.$$

When $G = 0$, the Riccati equation is rank preserving. It defines then a vector field on the sub-manifold of rank $r < n$ covariance matrices ($n = \dim x$ here). This sub-manifold admits the over-parameterization

$$(O, R) \mapsto ORO^T = P \leftrightsquigarrow \overbrace{\rule{0.5em}{4em}}^{O}\ \overbrace{\rule{1em}{1em}}^{R}\ \overbrace{\rule{4em}{1em}}^{O^T} = \overbrace{\rule{4em}{4em}}^{P}$$

where O belongs to the set of $n \times r$ orthogonal matrices ($O^T O = \mathbb{I}_r$) and R is $r \times r$, positive definite and symmetric.
Lift of dP/dt ($P = ORO^T$ solution of the Riccati equation with $G = 0$):

$$\frac{d}{dt}O = (\mathbb{I}_n - OO^T)AO, \quad \frac{d}{dt}R = O^T AOR + RO^T AO - RO^T C^T (HH^T)^{-1} COR.$$

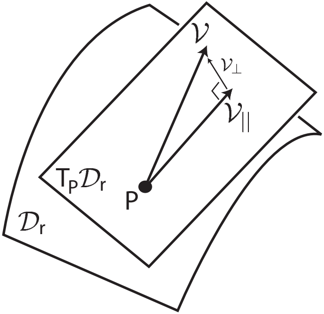
---

[1] S. Bonnabel and R. Sepulchre. *Matrix Information Geometry*, chapter The Geometry of Low-Rank Kalman Filters, pages 53–68. Springer, 2012. Preliminary version: arXiv:1203.4049v1.

Orthogonal projection on $\mathcal{D}_r$, the set of rank r symmetric matrices

$$\frac{d}{dt}P = \mathcal{V}(P) = AP + PA^T + GG^T - PC^T(HH^T)^{-1}CP$$

$P \geq 0$, $n \times n$ symmetric matrix
Select $r \leq n$ and consider the sub-set of symmetric matrices of rank r, i.e. a sub-manifold $\mathcal{D}_r$ of the vector space of symmetric matrices equipped with Frobenius Euclidian metric.
The approximate evolution is given by the orthogonal projection $\Pi_r^P(\mathcal{V}(P))$ of $\mathcal{V}(P)$ onto the tangent space $T_P\mathcal{D}_r$ at P to $\mathcal{D}_r$.



The low-rank approximation of Riccati differential equation:

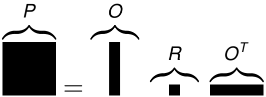
$$\text{for } P \in \mathcal{D}_r, \quad \frac{d}{dt}P = \mathcal{V}_{\|}(P) = \overbrace{\Pi_r^P\Big(AP + PA^T + GG^T - PC^T(HH^T)^{-1}CP\Big)}^{\text{vector field on } \mathcal{D}_r}.$$

Main issue: computation of $\mathcal{V}_{\|}(P)$ in "coordinates".

Projection and lift for rank-r covariance matrices

The sub-manifold $\mathcal{D}_r$ of covariance matrices P of rank $r < n$ is over-parameterized via

$$P = ORO^T \leftrightsquigarrow$$



where R is a $r \times r$ strictly positive positive matrix, O a $n \times r$ matrix with $O^T O = \mathbb{I}_r$.

Family of lifts for $\frac{d}{dt}P = \Pi_r^P \Big(AP + PA^T + GG^T - PC^T(HH^T)^{-1}CP \Big)$

$$\frac{d}{dt}O = \boldsymbol{\Omega} O + (\mathbb{I}_n - OO^T)\big((A - \boldsymbol{\Omega})O + GG^T O R^{-1}\big)$$

$$\frac{d}{dt}R = O^T(A - \boldsymbol{\Omega})OR + RO^T(A^T + \boldsymbol{\Omega})O$$
$$- RO^T C^T (HH^T)^{-1} COR + O^T GG^T O$$

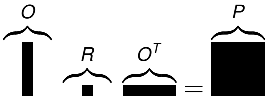
where the gage degree of freedom $\boldsymbol{\Omega}$ is any time varying $n \times n$ shew-symmetric matrix $\Omega^T = -\Omega$.

O remains orthogonal $O^T O = \mathbb{I}_r$. R remains symmetric > 0 and of rank r.

The computation of the lifted dynamics

Tangent map of the submersion:

$$(O, R) \mapsto ORO^T = P \leftrightsquigarrow$$



with the infinitesimal variations $\delta O = \omega O$ and $\delta R = \sigma$:

$$(\omega, \sigma) \mapsto [\omega, P] + O\sigma O^T$$

where ω is any $n \times n$ skew-symmetric matrix, σ is any $r \times r$ symmetric matrix.
A $n \times n$ symmetric matrix ρ in the tangent space at $P = ORO^T$ to $\mathcal{D}_r$ admits the parameterization $\rho = [\omega, P] + O\sigma O^T$.
The projection $\Pi_r^P\left(\frac{d}{dt}P\right)$ corresponds to the tangent vector ρ associated to $\boldsymbol{\omega}$ and $\boldsymbol{\sigma}$ minimizing

$$\operatorname{Tr}\left(\left(AP + PA^T + GG^T - PC^T(HH^T)^{-1}CP - [\boldsymbol{\omega}, P] - O\boldsymbol{\sigma}O^T\right)^2\right),$$

First order stationary conditions give $\boldsymbol{\omega}$ and $\boldsymbol{\sigma}$ as function of (O, R): the lifted evolution is given by $\frac{d}{dt}O = \boldsymbol{\omega}O$ and $\frac{d}{dt}R = \boldsymbol{\sigma}$ where the arbitrary skew-symmetric matrix Ω appears.

For $P = ORO^T$ lift of projected dynamics $\mathcal{V}_{||}(P) = \Pi_r^P(\mathcal{V}(P))$

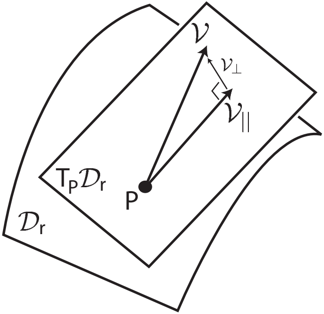
Data: $P \mapsto \mathcal{V}(P)$. For any skew-symmetric $n \times n$ matrix Ω, set

$$\begin{aligned}\Psi_\Omega(O,R) &= (\mathbb{I}_n - OO^T)\, \mathcal{V}\left(ORO^T\right)\, OR^{-1} + OO^T\Omega O\\ \Phi_\Omega(O,R) &= O^T\, \mathcal{V}\left(ORO^T\right)\, O + \left[R, O^T\Omega O\right].\end{aligned}$$

Assume that

$$\frac{d}{dt}O = \Psi_\Omega(O,R), \qquad \frac{d}{dt}R = \Phi_\Omega(O,R)$$

then $P = ORO^T$ satisfies automatically $\frac{d}{dt}P = \mathcal{V}_{||}(P)$.
Gage degree of freedom : arbitrary time-varying skew-symmetric $n \times n$ matrix Ω.



Moreover OO^T is the orthogonal projector on the image of $P = ORO^T$ with

$$\mathcal{V}_{||} + \mathcal{V}_\perp = \mathcal{V}, \quad \mathcal{V}_\perp = (\mathbb{I}_n - OO^T)\, \mathcal{V}\, (\mathbb{I}_n - OO^T).$$

Euler numerical scheme preserving positiveness of R

With $M = GG^T \geq 0$, $N = C^T(HH^T)^{-1}C \geq 0$ and gage $\Omega = 0$:

$$\frac{d}{dt}O = (\mathbb{I}_n - OO^T)\big(AO + MO\,R^{-1}\big)$$
$$\frac{d}{dt}R = O^TAOR + RO^TA^TO - RO^TNOR + O^TMO$$

O_k and R_k the numerical approximations of $O(k\delta t)$ and $R(k\delta t)$:

$$O_{k+1} = \Upsilon\Big(O_k + \delta t\ \big(\mathbb{I}_n - O_kO_k^t)(AO_k + MO_kR_k^{-1}\big)\Big)$$
$$R_{k+1} = \big(\mathbb{I}_r + \delta t\ W_k\big)R_k\big(\mathbb{I}_r + \delta t\ W_k^T\big) + \delta t\ O_k^TMO_k.$$

where

- $W_k = O_k^TAO_k - \frac{1}{2}R_kO_k^TNO_k$
- Υ stands for orthonormalization to ensure $O_{k+1}^TO_{k+1} = \mathbb{I}_r$.

1D wave equation with partial state measurement

- Consider the wave equation

$$\frac{\partial^2 h}{\partial t^2} = \frac{\partial^2 h}{\partial z^2} + \tfrac{1}{2}\eta(z,t) \quad z \in [0,1] \text{ with } h(0,t) = h(1,t)$$

  and $\eta(z,t)$ independent Wiener processes.

- The measure of h is incomplete

$$z \in [0, \tfrac{1}{2}], \quad y(z,t) = h(z,t) + \mu(z,t)$$

  with $\mu(z,t)$ independent Wiener processes.

Numerical simulations

- Discretization mesh $z \in \{\frac{1}{m}, \ldots, \frac{i}{m}, \ldots, \frac{m}{m}\}$
- The $m > 0$ $h(i/m, t) = h_i(t)$ obey to the second order ODE's

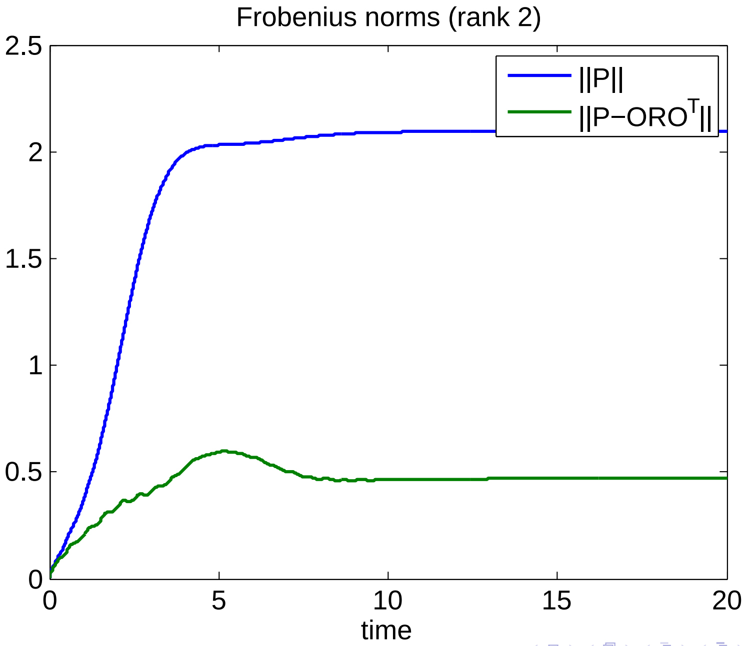
$$\frac{d^2}{dt^2} h_i(t) = m^2 (h_{i+1}(t) + h_{i-1}(t) - 2h_i(t)) + \eta_i(t), \quad i \in \{1, \ldots, m\}$$

  with $y_i(t) = h_i(t) + \mu_i(t)$ for $i \in \{1, \ldots, m/2\}$.
- State $x = (h_i, \dot{h}_i)_{1 \leq i \leq n}$ of dimension $n = 2m$: $\frac{d}{dt} x = Ax + G\eta$, $y = Cx + \mu$.
- Initialization Riccati: $P(0) = 0_{n \times n}$, $O(0) = \mathbb{I}_{n \times r}$, $R = \mathbb{I}_{r \times r}/100$.
- Observer convergence:

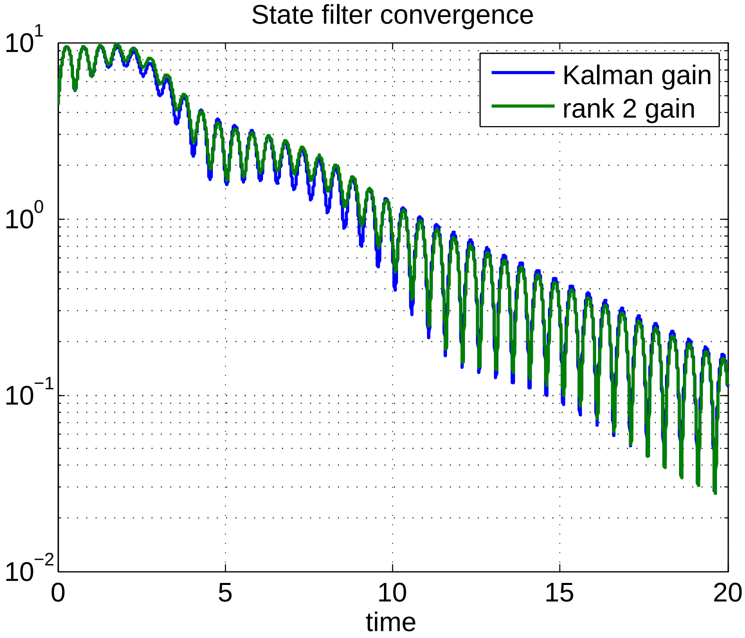
$$\frac{d}{dt} \widetilde{x} = (A - LC)\widetilde{x} \quad \text{with } L = PC^T (HH^T)^{-1} \text{ or } ORO^T C^T (HH^T)^{-1}$$

  with P, O and R solutions of the Riccati differential equation and its rank r orthogonal projection with gage $\Omega = 0$.
  Initialization: $\widetilde{x}(0)$ corresponds to $h(z, 0) = \sin(\pi z)$ and $\frac{\partial h}{\partial t}(z, 0) = \sin(\pi z)$.

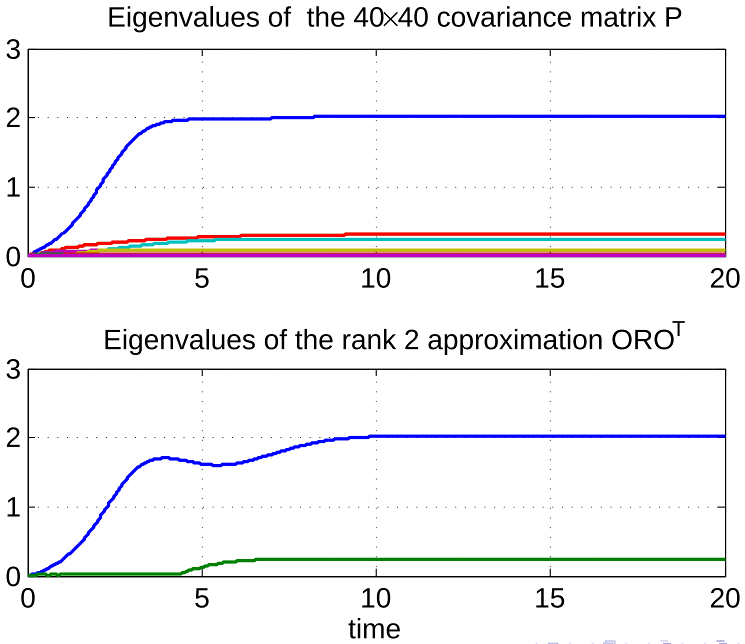
Difference between P and ORO^T (rank $r = 2$ with $m = 20$)



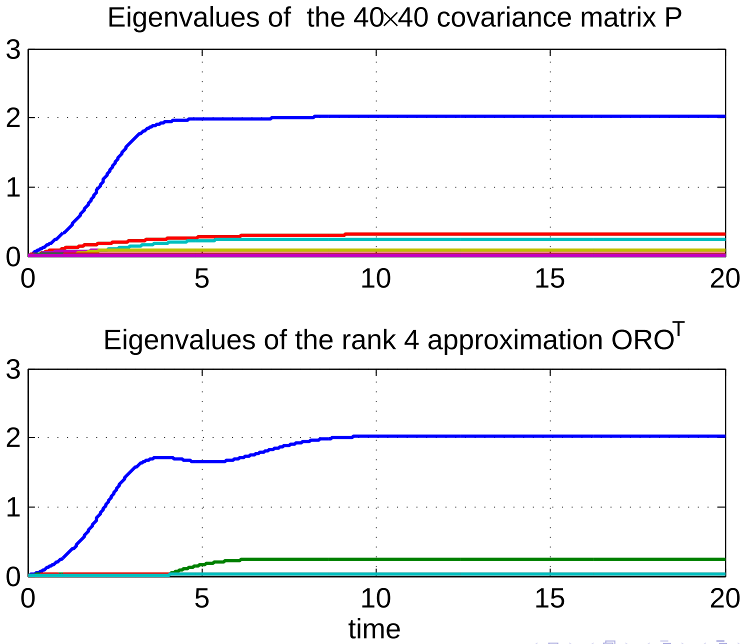
Convergence rate of $A - LC$ (rank $r = 2$ with $m = 20$)



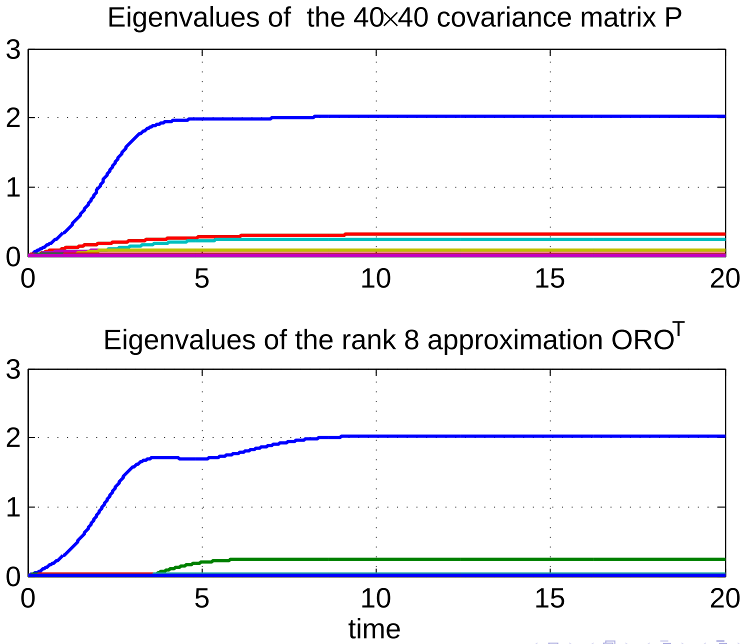
Spectra of P and ORO^T (rank $r = 2$ with $m = 20$)



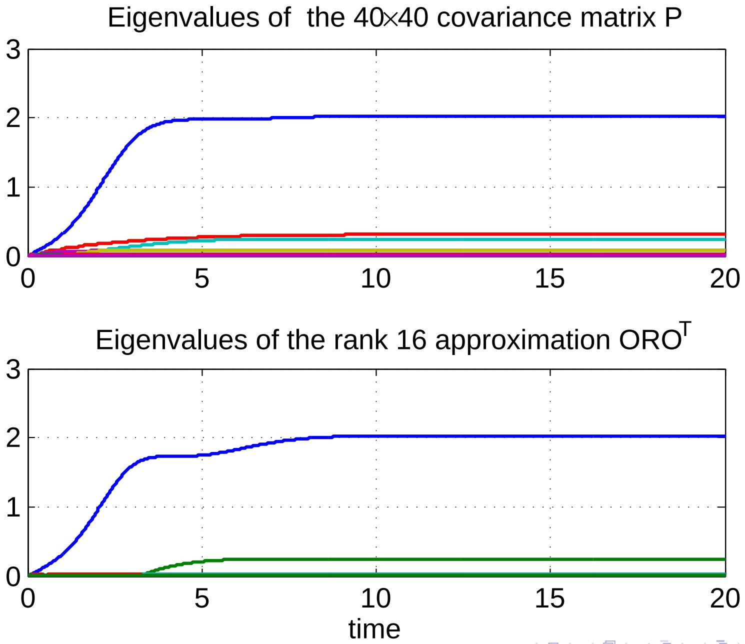
Spectra of P and ORO^T (rank $r = 4$ with $m = 20$)



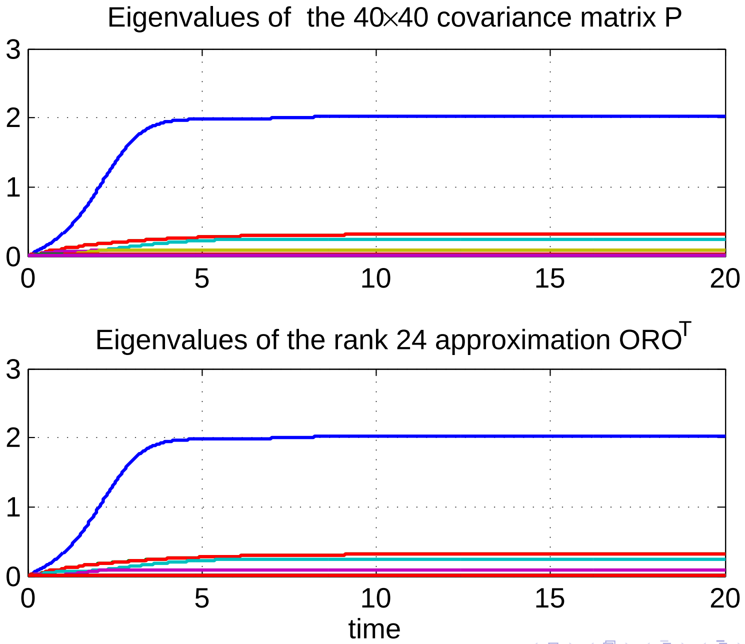
Spectra of P and ORO^T (rank $r = 8$ with $m = 20$)



Spectra of P and ORO^T (rank $r = 16$ with $m = 20$)



Spectra of P and ORO^T (rank $r = 24$ with $m = 20$)



Concluding remarks

- Use a gage $\Omega \neq 0$. Initialization of O and R...
- Possible parallelization of products like AO, ...
- Extension to discrete-time Kalman filtering: adaptation to SEEK filters ?

  Orthogonal projection of P on $\mathcal{D}_r$ is a non-linear operation with possible ambiguity when the spectrum of P is degenerate ...
- Low-rank approximation $\rho = U\sigma U^\dagger$ for Lindbald differential equation describing open-quantum systems:

$$\frac{d}{dt}\rho = -i[H, \rho] + \sum_\nu L_\nu \rho L_\nu^\dagger - \tfrac{1}{2}(L_\nu^\dagger L_\nu \rho + \rho L_\nu^\dagger L_\nu)$$

  arXiv preprint http://arxiv.org/abs/1207.4580
  with rank 16 approximation of a density matrix ρ of size 15000×15000.
  (oscillation revivals of 50 two-level atoms resonantly coupled to a coherent field with an average of 200 photons).