Low-rank and sparse methods for high-dimensional approximation and model order reduction

Lecture 9

Geometry of tensor manifolds and applications

Introduction

In this lecture, we describe the geometric structures of manifolds of tree-based low-rank tensors and present two important applications:

- the optimization in low-rank tensors manifolds,
- the dynamical low-rank approximation for dynamical systems in tensor spaces.

Outline

Outline

1. Geometry of tensor manifolds
2. Optimization in low-rank tensor manifolds
3. Dynamical low-rank approximation

Geometry of low-rank tensor manifolds

Subsets of tensors with fixed tree-based rank r admits a multilinear parametrization

$$\mathcal{M}_r = \{u = F(p_1, \dots, p_M) : p_\nu \in P_\nu, 1 \leq \nu \leq M\}$$

where the p_ν are parameters in some vector spaces P_ν, and where F is a multilinear map.

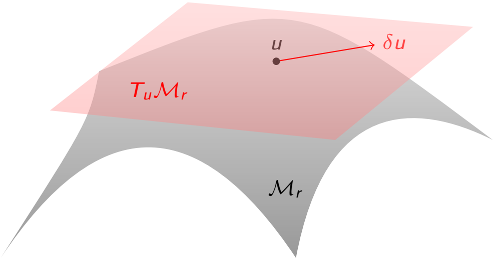
The map F is surjective but usually not injective, which means that a tensor $u = F(p_1, \dots, p_M)$ has no unique representation.

Understanding the equivalence between the representations of a tensor is the key for understanding the geometry of the manifold $\mathcal{M}_r$.

Geometry of low-rank tensor manifolds

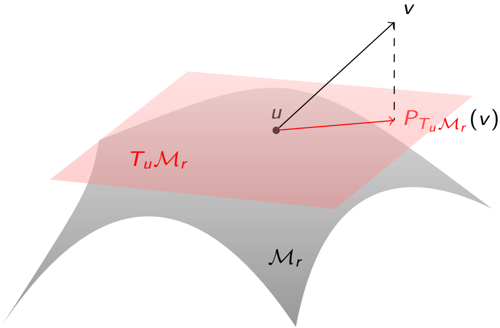
The characterization of the tangent space of $\mathcal{M}_r$ at u, denoted $T_u\mathcal{M}_r$, is of particular importance in applications. Due to the multilinearity of F, elements of $T_u\mathcal{M}_r$, called tangent vectors, admit a representation of the form

$$\delta u = F(\delta p_1, p_2, \dots, p_M) + \dots + F(p_1, p_2, \dots, \delta p_M), \quad \delta p_1 \in P_1, \dots, \delta p_M \in P_M.$$



Geometry of low-rank tensor manifolds

The representation of a tangent vector δu is usually made unique by imposing some conditions on the parameters δp_ν, which allows to define a projection operator $P_{T_u\mathcal{M}_r}$ onto the tangent space $T_u\mathcal{M}_r$.



Manifold of low-rank matrices

Let $\mathcal{M}_r$ be the set of matrices in $\mathbb{R}^{n\times m}$ with a fixed rank r

$$\mathcal{M}_r = \{u \in \mathbb{R}^{n\times m} : \text{rank}(u) = r\}$$

A matrix $u \in \mathbb{R}^{n\times m}$ with rank r can be written as

$$u = UV^T$$

where U and V belongs to the non compact Stiefel manifold $\mathbb{R}_*^{n\times r}$ of matrices with full rank. Also, it can be written as

$$u = UCV^T$$

where U (resp. V) belongs to the Stiefel manifold $St(n, r)$ (resp. $St(m, r)$) of orthogonal matrices, and where C belongs to the linear group $GL(r)$ of invertible matrices.

This yields two different parametrizations of $\mathcal{M}_r$ (among others), each of which inducing different looks at the geometry of $\mathcal{M}_r$.

Manifold of low-rank matrices

In the following, we consider the parametrization

$$\mathcal{M}_r = \{u = F(C, U, V) : U \in St(n, r), V \in St(m, r), C \in GL(r)\}.$$

where

$$F : St(r, n) \times St(r, m) \times GL(r) \to \mathcal{M}_r$$

is a multilinear map defined by

$$F(C, U, V) \mapsto UCV^T.$$

The natural extension of this parametrization to higher-order tensors is the Tucker format.

A tensor $u = F(C, U, V) \in \mathcal{M}_r$ admits infinitely many equivalent representations of the form

$$u = F(C, U, V) = F(ACB, UA^{-1}, VB^{-T}) \quad \text{for } A, B \in GL(r). \tag{1}$$

The relation (1) defines an equivalence relation between parameters which suggests that the relevant parameters are not the orthogonal matrices U and V but the subspaces $\mathcal{U}$ and $\mathcal{V}$ spanned by their columns.

Manifold of low-rank matrices

The tangent space to $\mathcal{M}_r$ at point $u = F(C, U, V)$, denoted by $T_u\mathcal{M}_r$, is the set of elements δu of the form

$$\delta u = F(\delta C, U, V) + F(C, \delta U, V) + F(C, U, \delta V)$$

where $\delta C \in \mathbb{R}^{r\times r}$, $\delta U \in \mathbb{R}^{n\times r}$, $\delta V \in \mathbb{R}^{m\times r}$.

The representation of elements of $T_u\mathcal{M}_r$ can be made unique by imposing δU and δV to be orthogonal to U and V respectively. This yields the characterization of the tangent space

$$T_u\mathcal{M}_r = \Big\{ U\delta C V^T + \delta U C V^T + UC\delta V^T : \\ \delta C \in \mathbb{R}^{r\times r}, \delta U \in \mathbb{R}^{n\times r}, \delta V \in \mathbb{R}^{m\times r}, \delta U^T U = 0, \delta V^T V = 0 \Big\}.$$

The orthogonal projection operator $P_{T_u\mathcal{M}_r}$ onto $T_u\mathcal{M}_r$ is defined by

$$P_{T_u\mathcal{M}_r} = P_{\mathcal{U}} \otimes P_{\mathcal{V}} + P_{\mathcal{U}^\perp} \otimes P_{\mathcal{V}} + P_{\mathcal{U}} \otimes P_{\mathcal{V}^\perp}$$

Manifold of tensors with fixed tree-based rank

The previous considerations naturally extend to the case of the manifold $\mathcal{M}_r$ of tensors in $\mathbb{R}^{n_1} \otimes \ldots \otimes \mathbb{R}^{n_d}$ with Tucker rank $(r_1, \ldots, r_d)$, which admits the parametrization

$$\mathcal{M}_r = \left\{ u = F(C, U_1, \ldots, U_d) : C \in \mathbb{R}_*^{r_1 \times \ldots \times r_d}, U_\nu \in St(r_\nu, n_\nu) \right\},$$

where $\mathbb{R}_*^{r_1 \times \ldots \times r_d}$ is the set of order-d tensors C such that all matricisations $\mathcal{M}_\nu(C)$ have full rank r_ν.

The tangent space to $\mathcal{M}_r$ at $u = F(C, U_1, \ldots, U_d)$ is defined by

$$T_u\mathcal{M}_r = \{ F(\delta C, U_1, \ldots, U_d) + F(C, \delta U_1, \ldots, U_d) + \ldots + F(C, U_1, \ldots, \delta U_d) : \\ \delta C \in \mathbb{R}^{r_1 \times \ldots \times r_d}, \delta U_\nu \in \mathbb{R}^{n_\nu \times r_\nu}, {\delta U_\nu}^T U_\nu = 0 \right\}.$$

and the orthogonal projection onto the tangent space is defined by

$$P_{T_u\mathcal{M}_r} = P_{\mathcal{U}_1} \otimes \ldots \otimes P_{\mathcal{U}_d} + P_{\mathcal{U}_1^\perp} \otimes P_{\mathcal{U}_2} \otimes \ldots \otimes P_{\mathcal{U}_d} + P_{\mathcal{U}_1} \otimes \ldots \otimes P_{\mathcal{U}_{d-1}} \otimes P_{\mathcal{U}_d^\perp}.$$

Manifold of tensors with fixed tree-based rank

The extension to more general subsets of tensors $\mathcal{M}_r$ with fixed tree-based tensor formats, although more subtle, is also possible.

$\mathcal{M}_r$ is proved to be a smooth and immersed submanifold of the tensor space, for a finite dimensional tensor space. This is also true, under some technical assumptions, for infinite-dimensional tensor Banach spaces.

Note that $\mathcal{M}_r$ can in fact be identified with a fiber bundle (which comes from the product structure of the parameter set). The tangent space can be split into vertical and horizontal tangent spaces. This structure has interesting consequences in optimization and dynamical low-rank approximation (not discussed here).

Outline

Optimization in low-rank tensor manifolds

Let us consider the optimization problem

$$\min_{v \in \mathcal{M}_r} \mathcal{J}(v)$$

where $\mathcal{M}_r$ is a subset of tensors with fixed tree-based rank, and assume that it admits a solution u.

Optimization in low-rank tensor manifolds

A necessary (but not sufficient) condition for optimality is

$$\langle \nabla \mathcal{J}(u), \delta u \rangle = 0, \quad \forall \delta u \in T_u \mathcal{M}_r.$$

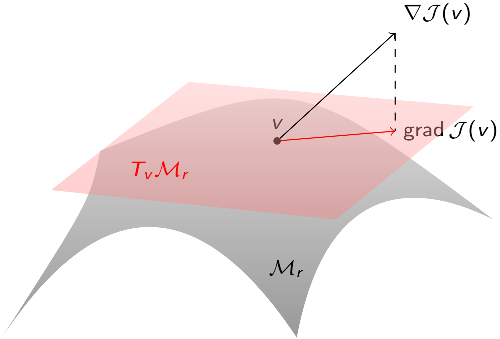
This is equivalent to

$$\operatorname{grad} \mathcal{J}(u) = 0,$$

where $\operatorname{grad} \mathcal{J}$ is the Riemannian gradient of $\mathcal{J}$ on $\mathcal{M}_r$ such that

$$\operatorname{grad} \mathcal{J}(v) = P_{T_v \mathcal{M}_r} \nabla \mathcal{J}(v),$$

where $P_{T_v \mathcal{M}_r}$ is the orthogonal projection onto the tangent space $T_v \mathcal{M}_r$ of $\mathcal{M}_r$ at v.

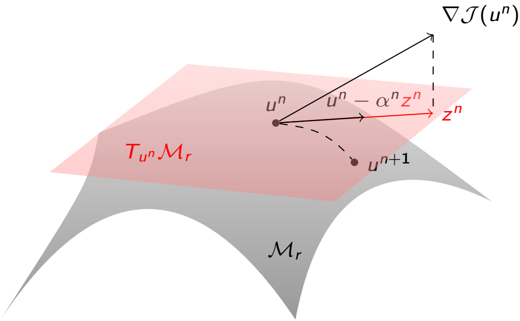


Gradient algorithms

Gradient algorithms in the Riemannian manifold $\mathcal{M}_r$ consist in computing a sequence u^n in $\mathcal{M}_r$ defined by

$$z^n = \text{grad}\, \mathcal{J}(u^n) = P_{T_{u^n}\mathcal{M}_r} \nabla \mathcal{J}(u^n),$$
$$u^{n+1} = R_{u^n}(u^n - \alpha^n z^n),$$

where $R_{u^n} : T_{u^n}\mathcal{M}_r \to \mathcal{M}_r$ maps an element of the tangent space $T_{u^n}\mathcal{M}_r$ onto $\mathcal{M}_r$.



It can be seen as a projected gradient algorithm with a nonlinear projection onto $\mathcal{M}_r$:

$$u^{n+1} = \Pi(u^n - \alpha^n \nabla \mathcal{J}(u^n)), \quad \Pi = R_{u^n} \circ P_{T_{u^n}\mathcal{M}_r} : V \to \mathcal{M}_r.$$

Computing projections

Assume that $\nabla\mathcal{J}(u^n)$ has a representation in low-rank format, e.g. when

$$\nabla\mathcal{J}(u^n) = Au^n - b$$

where A and b have low-rank representations.

The projection $P_{T_{u^n}\mathcal{M}_r}$ onto the tangent space $T_{u^n}\mathcal{M}_r$ of $\mathcal{M}_r$ also have a low-rank tensor structure.

Therefore, computing $z^n = P_{T_{u^n}\mathcal{M}_r}\nabla\mathcal{J}(u^n)$ requires standard tensor algebra (multiplication of a low-rank operator by a low-rank vector), and z^n has again a low-rank representation.

Retractions

For $u \in \mathcal{M}_r$ and $v \in T_u\mathcal{M}_r$, a retraction R_u maps smoothly $u+v$ to a point $R_u(v)$ in $\mathcal{M}_r$, such that

$$\|u + v - R_u(v)\| = o(\|v\|).$$

For immersed submanifolds, a particular retraction map is the metric projection which associates to v an element $R_u(v)$ of best approximation of $u+v$ in $\mathcal{M}_r$, such that

$$\|u + v - R_u(v)\| = \min_{w \in \mathcal{M}_r} \|u + v - w\|.$$

For the case of matrices (or order-two tensors), the metric projection corresponds to a truncated singular value decomposition (which is unique if v is sufficiently small).

For higher-order tensors, this best approximation problem can not be solved. A practical choice consists in taking for $R_u(v)$ the truncated higher-order singular value decomposition of $u+v$, which is such that

$$\|u + v - R_u(v)\| \leq C(d) \min_{w \in \mathcal{M}_r} \|u + v - w\|.$$

Steepest descent algorithms

A steepest descent algorithm is such that

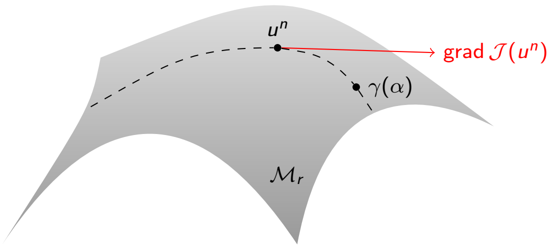
$$u^{n+1} = R_{u^n}(u^n - \alpha^n \,\mathrm{grad}\, \mathcal{J}(u^n)),$$

where α^n is the solution of the minimization problem

$$\min_{\alpha \in \mathbb{R}} \mathcal{J}(\gamma(\alpha)),$$

with

$$\gamma(\alpha) = R_{u^n}(u^n - \alpha \,\mathrm{grad}\, \mathcal{J}(u^n)).$$



Outline

Dynamical systems in tensor spaces

Consider a dynamical system

$$\dot{u}(t) = F(u(t), t),$$
$$u(0) = u_0,$$

defined on a tensor Hilbert space V, where

$$F : V \times \mathbb{R}^+ \to V.$$

Dynamical low-rank approximation

Assuming that the solution $u(t)$ at each time admits a good approximation in a certain low-rank manifold $\mathcal{M}_r$, the problem is then to define a reduced dynamical system in $\mathcal{M}_r$ which produces a dynamical low-rank approximation of $u(t)$.

Dynamical low-rank approximation

Example 1 (Time-dependent parameter-dependent equation)

$$\frac{\partial u}{\partial t} - \nabla \cdot (\kappa(\xi)\nabla u) = f(\xi), \quad u \in H^1(\Omega) \otimes L^2_\mu(\Xi)$$

where ξ are parameters, possibly random. A dynamical low-rank approximation of u is of the form

$$u_r(t,x,\xi) = \sum_{i=1}^{r} v_i(t,x)\lambda_i(t,\xi).$$

Example 2 (Time-dependent PDE defined on a high-dimensional product domain)

$$\frac{\partial u}{\partial t} + \sum_{i=1}^{d} \frac{\partial}{\partial x_i}(a_i u) - \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2}{\partial x_i x_j}(b_{ij} u) = 0, \quad u \in H^1(\mathcal{X}_1) \otimes \ldots \otimes H^1(\mathcal{X}_d),$$

A dynamical low-rank approximation of u in tensor-train format is of the form

$$u_r(t,x_1,\ldots,x_d) = \sum_{k_1=1}^{r_1} \ldots \sum_{k_{d-1}=1}^{r_{d-1}} v^{(1)}(t,x_1,k_1)v^{(2)}(t,x_2,k_1,k_2)\ldots v^{(d)}(t,x_d,k_{d-1}).$$

Dirac-Frenkel variational principle

Assuming that $u_0 \in \mathcal{M}_r$, the Dirac-Frenkel variational principle defines a reduced dynamical system with a solution $t \mapsto u_r(t) \in \mathcal{M}_r$ such that $u_r(0) = u_0$ and

$$\|\dot{u}_r(t) - F(u_r(t), t)\| = \min_{\dot{v} \in T_{u_r(t)}\mathcal{M}_r} \|\dot{v} - F(u_r(t), t)\|,$$

or equivalently

$$\langle \dot{u}_r(t) - F(u_r(t), t), \dot{v} \rangle = 0 \quad \forall \dot{v} \in T_{u_r(t)}\mathcal{M}_r,$$

which defines $\dot{u}_r(t)$ as the projection of the flux on the tangent space to $\mathcal{M}_r$ at $u_r(t)$, i.e.

$$\dot{u}_r(t) = P_{T_{u_r(t)}\mathcal{M}_r} F(u_r(t), t).$$