GEOMETRIC SUBSPACE UPDATES WITH APPLICATIONS TO ONLINE ADAPTIVE NONLINEAR MODEL REDUCTION*

RALF ZIMMERMANN†, BENJAMIN PEHERSTORFER‡, AND KAREN WILLCOX§

Abstract. In many scientific applications, including model reduction and image processing, subspaces are used as ansatz spaces for the low-dimensional approximation and reconstruction of the state vectors of interest. We introduce a procedure for adapting an existing subspace based on information from the least-squares problem that underlies the approximation problem of interest such that the associated least-squares residual vanishes exactly. The method builds on a Riemannian optimization procedure on the Grassmann manifold of low-dimensional subspaces, namely the Grassmannian Rank-One Update Subspace Estimation (GROUSE). We establish for GROUSE a closed-form expression for the residual function along the geodesic descent direction. Specific applications of subspace adaptation are discussed in the context of image processing and model reduction of nonlinear partial differential equation systems.

Key words. online adaptive model reduction, dimension reduction, Grassmann manifold, Grassmannian Rank-One Update Subspace Estimation (GROUSE), discrete empirical interpolation method (DEIM), gappy proper orthogonal decomposition (POD), masked projection, rank-one updates, optimization on manifolds, subspace fitting, least-squares, image processing

AMS subject classifications. 65F30, 65F20, 49M15, 65K10, 37M99

DOI. 10.1137/17M1123286

1. Introduction. Dimension reduction techniques play an important role in the application of computational methods—identifying inherent low-dimensional structure in the problem at hand can often lead to significant reductions in computational complexity. Consider a set of state vectors embedded in the n-dimensional Euclidean space \( \mathbb{R}^n \), \( n \in \mathbb{N} \). The goal of dimension reduction is to restrict the space of state vector candidates to a subspace of \( \mathbb{R}^p \) of low dimension \( p \ll n \). In doing so, the \( n \)-degree-of-freedom problem of computing full-scale state vectors is replaced by the task of determining the \( p \) coefficients of a basis expansion in the reduced subspace. If, for example, the state vectors are solutions of a computational model, then this dimension reduction underlies the derivation of a projection-based reduced model. As another example, the state vectors might represent experimental data or other system samples such as representations of an image. In those cases, the dimension reduction seeks an efficient compression of the data and a low-dimensional subspace in which to reconstruct unknown states. When \( n \) is large, dimension reduction often leads to a tremendous reduction in computational complexity; however, acceptable accuracy is only retained if the full state vectors can be approximated well in the \( p \)-dimensional

*Received by the editors March 29, 2017; accepted for publication (in revised form) by P.-A. Absil November 20, 2017; published electronically February 15, 2018.
http://www.siam.org/journals/simax/39-1/M112328.html

Funding: The work of the first author was funded via a research fellowship awarded by the German Research Foundation (DFG), grant Zi 1250/2-1. The work of the second and third authors was supported by the U.S. Department of Energy, awards DE-FG02-08ER2585 and DE-SC0009297, as part of the DiaMonD Multifaceted Mathematics Integrated Capability Center, and the AFOSR MURI grant FA9550-15-1-0038.

†Department of Mathematics and Computer Science, SDU Odense, Odense DK-5230, Denmark, and Department of Aeronautics & Astronautics, MIT, Cambridge, MA 02139 (zimmermann@imada.sdu.dk).
‡Department of Mechanical Engineering, University of Wisconsin-Madison, Madison, WI 53706 (peherstorfer@wisc.edu).
§Department of Aeronautics & Astronautics, MIT, Cambridge, MA 02139 (kwillcox@mit.edu).
subspace. Thus, the identification and numerical representation of subspaces plays a
critical role.

In classical projection-based model reduction, the reduced subspace is determined
once in a so-called offline phase. Subsequently, it stays fixed while the reduced model
is evaluated during the so-called online phase. Online adaptive model reduction breaks
this division, and modifies the subspace during the evaluation process to better meet
the current conditions for the reduced state vector prediction.

Online subspace adaptation can be approached from a geometric perspective:
The set of all subspaces $U \subseteq \mathbb{R}^n$ of a certain fixed dimension $p$ forms the Grassmann
manifold $[2]$. Subspaces are spatial locations on this manifold and are represented in
numerical schemes by column-orthogonal matrices in $\mathbb{R}^{n \times p}$. One-parameter subspace
modifications correspond to curves on the Grassmannian.

In the special case, where the subspace adaptation is based on a linear least-
squares residual function, the Grassmannian Rank-One Update Subspace Estimation
(GROUSE, [8]) applies: When approximating an unsampled state vector in the sub-
space $U$ based on partial information, the associated least-squares residual is then
related to a velocity vector of a geodesic curve on the Grassmannian. GROUSE shows
that this geodesic curve corresponds to a matrix curve of rank-one modifications on
the underlying column-orthogonal matrices that act as subspace representatives.

**Main contributions.** We show that the GROUSE geodesic of rank-one updates
crosses a subspace $U^*$ that allows for an exact representation of the given partial
information. Mathematically, this is a nonlinear root-finding problem on the Grass-
mann manifold. We derive a closed-form expression for the residual with respect to
the partial information along the GROUSE geodesic. In particular, this allows us to
read off the root, but it may be of potential use in general when analyzing GROUSE
with other step size schemes. As an auxiliary, we establish a general formula for the
rank-one update of orthogonal projectors. Moreover, we generalize the method to
subspace adaptation based on general least-squares systems and to the adaptation of
a subspace of the subspace in question.

In the results section, we demonstrate that the proposed method applies in com-
bination with the following well-established dimension reduction techniques: gappy
proper orthogonal decomposition (gappy POD, [27, 18]) and discrete empirical inter-
polation method (DEIM, [22]). More precisely, we consider an application to gappy
POD image processing, and we combine the subspace adaptation with the DEIM to
construct an adaptive reduced model for the time-dependent nonlinear FitzHugh–
Nagumo partial differential equation system, which models the electrical activity in a
neuron. In contrast to the standard use case in the GROUSE literature [8, 49], our
focus is not on estimating a subspace from scratch based on potentially noisy data but
to adapt a given subspace of valid approximations based on incomplete but noise-free
observations. In the DEIM setting, it is not the final subspace that is of main interest
but rather the enhanced approximation capabilities after each adaptation.

**Context and related work.** The Grassmann manifold can be represented as a ma-
trix manifold. For comprehensive background information on optimization on matrix
manifolds, we refer the reader to [2] and its extensive bibliography. Matrix manifolds
appear frequently in image processing and computer vision [35], where they often take
the form of subspace identification problems. A related field of application is low-rank
matrix factorizations, which arise in data analysis problems of various kinds, among
them matrix completion [8, 15]. The GROUSE method was introduced in [8] as a tool
for both subspace identification from incomplete and/or noisy data and the matrix
completion problem and was further developed and analyzed in [10, 31, 48, 49].
A recent survey of model reduction methods for parametric systems is [14]. Most online adaptive model reduction techniques rely on precomputed quantities that restrict the way the reduced space can be changed online. One example is parametric model reduction based on the interpolation of reduced models, where reduced operators are interpolated on matrix manifolds [3, 23, 38, 4, 36, 50]. There are also dictionary approaches [30, 34] that construct a reduced space online from a subset of a large number of precomputed basis vectors, and localized reduced modeling techniques [5, 40, 26, 24] that select online one of several precomputed reduced models.

In contrast, we are interested here in online adaptive model reduction methods that derive updates to the reduced model with information that is obtained from the full model in the online phase; thus, the adaptation uses information that is unavailable in the offline phase. There are several approaches that generate new data from the full model in the online phase, or derive new reduced basis vectors with an $h$-refinement [21] based on an adjoint model of the full model, and then rebuild the reduced model [37, 39, 44, 45]; however, this is often computationally expensive. An efficient online adaptation that uses new data online was presented in [46, 6] for localized reduced models. A reference state is subtracted from the snapshots of localized reduced models. It is shown that this corresponds to a rank-one update of the reduced space corresponding to the localized reduced models; however, this is only a limited form of adapting a reduced model because each snapshot receives the same change. In [42, 41], dynamic reduced models are introduced that adapt to changes in the full model without requiring access to the high-fidelity operators; however, the approach is limited to linear problems and to problems where high-resolution sensor information is available that provides approximations of the full state vectors. For nonlinear problems, an adaptive DEIM was presented in [43], which derives low-rank updates to the DEIM basis from sparse data of nonlinear terms. In this paper we draw on the theory of Grassman manifolds and subspace updates to introduce a more flexible method for adaptive model reduction that applies to nonlinear problems and reproduces the inputted sparse data exactly.

**Notation and preliminaries.** The $(p \times p)$-identity matrix is denoted by $I_p \in \mathbb{R}^{p \times p}$. If the dimension is clear, we will simply write $I$. The $(p \times p)$-orthogonal group, i.e., the set of all square orthogonal matrices, is denoted by

$$O_p = \{ R \in \mathbb{R}^{p \times p} | R^T R = R R^T = I_p \}.$$ 

For a matrix $U \in \mathbb{R}^{n \times p}$, the subspace spanned by the columns of $U$ is denoted by $U := \text{colspan}(U) := \{ U \alpha \in \mathbb{R}^n | \alpha \in \mathbb{R}^p \} \subset \mathbb{R}^n$. The set of all $p$-dimensional subspaces $U \subset \mathbb{R}^n$ forms the Grassmann manifold

$$Gr(n,p) := \{ U \subset \mathbb{R}^n | \text{subspace, } \dim(U) = p \}.$$ 

The Stiefel manifold is the compact matrix manifold of all column-orthogonal rectangular matrices

$$St(n,p) := \{ U \in \mathbb{R}^{n \times p} | U^T U = I_p \}.$$ 

The Grassmann manifold can be realized as a quotient manifold of the Stiefel manifold

$$Gr(n,p) = St(n,p)/O_p = \{ [U] | U \in St(n,p) \},$$ 

where $[U] = \{ UR | R \in O_p \}$ is the orbit, or equivalence class of $U$ under actions of the orthogonal group. Hence, by definition, two matrices $U, \hat{U} \in St(n,p)$ are in the same
A matrix \( U \in \text{St}(n,p) \) is called a matrix representative of a subspace \( U \in \text{Gr}(n,p) \) if \( U = \text{colspan}(U) \). We will also consider the orbit \([U]\) and the subspace \( U = \text{colspan}(U) \) as the same object. As in [25], we will make use throughout of the quotient representation (1) of the Grassmann manifold with matrices in \( \text{St}(n,p) \) acting as representatives in numerical computations. From the manifold perspective, each \( p \)-dimensional subspace of \( \mathbb{R}^n \) is a single point on \( \text{Gr}(n,p) \).

For a rectangular, full column rank matrix \( X \in \mathbb{R}^{n \times p} \), the orthogonal projection onto the column span of \( X \) is

\[
\Pi_X : \mathbb{R}^n \to \text{colspan} X, \quad y \mapsto X(X^TX)^{-1}X^Ty.
\]

We will consider special orthogonal projectors associated with the Cartesian coordinate directions. Let \( e_j \in \mathbb{R}^n \) denote the \( j \)th canonical unit vector, \( j = 1, \ldots, n \). Given a subset of \( m \in \mathbb{N} \) indices \( J = \{j_1, \ldots, j_m\} \subset \{1, \ldots, n\} \), the (column-orthogonal) matrix \( P = (e_{j_1}, \ldots, e_{j_m}) \in \{0, 1\}^{n \times m} \) is called the mask matrix corresponding to the index set \( J \). Left-multiplication of a vector with the transpose of \( P \) realizes the projection onto the selected components in the same order as listed in \( J \), i.e., \( P^Ty = (y_{j_1}, \ldots, y_{j_m})^T \in \mathbb{R}^m \) for all \( y \in \mathbb{R}^n \). The matrix \( PP^T \) is the canonical orthogonal projection onto the coordinate axes \( j_1, \ldots, j_m \).

Throughout, whenever a mask matrix \( P \in \mathbb{R}^{n \times m} \) is applied to a subspace representative \( U \in \text{St}(n,p) \), we assume that \( m > p \) and that the matrix of selected rows \( P^TU \in \mathbb{R}^{m \times p} \) has full column rank \( p \).

**Organization.** Section 2 recaps the GROUSE approach and transfers the idea of the geometric subspace adaptation to the context of model reduction. It also reviews the essentials on the numerical treatment of Grassmann manifolds. Section 3 presents the core methodological contributions of this paper, where we derive a closed-form of the Grassmann rank-one update that solves the underlying least-squares residual equation exactly. Example applications in the context of adaptive model reduction and image processing are presented in section 4, and section 5 concludes the paper.

### 2. Problem statement

In this section, we first summarize GROUSE following [8]. We then develop the connection between the theory of GROUSE and the task of adapting a low-dimensional subspace for model reduction. Lastly, we discuss relevant concepts in the numerical treatment of Grassmann manifolds.

#### 2.1. GROUSE

Let \( P = (e_{j_1}, \ldots, e_{j_m}) \in \{0, 1\}^{n \times m} \) be a mask matrix, let \( U_0 \subset \mathbb{R}^n \) be a \( p \)-dimensional subspace with matrix representation \( U_0 = [U_0] \), \( U_0 \in \text{St}(n,p) \), and let \( b \in \mathbb{R}^m \) be a given data vector, \( p < m < n \). GROUSE considers the masked least-squares problem

\[
y(U_0) := \arg \min_{\hat{y} \in U_0} \|P^T\hat{y} - b\|_2^2,
\]

which features the (subspace dependent) unique solution

\[
y(U_0) = U_0\alpha(U_0) \in \mathbb{R}^n, \quad \alpha(U_0) = (U_0^T P P^T U_0)^{-1}U_0^T P b \in \mathbb{R}^p.
\]

The corresponding residual vector \( r(U_0) := b - P^T y(U_0) \) is, in general, nonzero. For a fixed mask matrix \( P \) and a fixed right-hand side \( b \), the residual vector is associated
with a differentiable function on $Gr(n, p)$, the residual norm function

$$
(5) \quad F_{P, b} : Gr(n, p) \to \mathbb{R}, \quad \mathcal{U} \mapsto \|r(\mathcal{U})\|_2^2 = b^T b - b^T P^T U (U^T P P^T U)^{-1} U^T P b;
$$

see [8, eqs. (2) and (3)]. (The matrix $U$ in the definition of $F_{P, b}$ can be any representative $U \in St(n, p)$ of the subspace $\mathcal{U}$; see (1). The subscripts $P, b$ will be dropped when clear from the context.) Given a sequence of incomplete observations in the form of data vectors $b_s \in \mathbb{R}^m, s = 1, 2, \ldots$, with corresponding mask matrices $P_s$, GROUSE adapts the initial subspace such that the objective

$$
(6) \quad \mathcal{U} \mapsto \sum_{s=1}^{\infty} F_{P_s, b_s}(\mathcal{U}) = \sum_{s=1}^{\infty} \|P_s^T y(\mathcal{U}) - b_s\|^2
$$

is minimized; see [8, eq. (5)].

The GROUSE algorithm works sequentially by addressing one data vector $b_s$ at a time. It performs a step along the geodesic line on $Gr(n, p)$ [25, section 2.5.1] along this descent direction leads to

$$
(7) \quad t \mapsto U_0(t) = U_0 + \left( (\cos(t\sigma_1) - 1) U_0 v + \sin(t\sigma_1) \frac{P r}{\|r\|} \right) v^T =: U_0 + \hat{\nu}(t) v^T;
$$

see [8, eqs. (11) and (12)]. At each iteration $s = 1, 2, \ldots$, the GROUSE algorithm [8, Alg. 1] chooses a step size $t = \eta_s$ and replaces the previous subspace representative $U_{s-1}$ by $U_s = U_{s-1}(\eta_s)$ according to (7). Local and global convergence results are given in [9, 48, 49].

### 2.2. Subspace adaptation and model reduction.

We consider here projection-based model reduction methods. These methods make use of a subspace $\mathcal{U}_0 \subset \mathbb{R}^n$ of comparatively low dimension $\dim(\mathcal{U}_0) = p \ll n$ that is assumed to contain the essential information about a set $\mathcal{X} \subset \mathbb{R}^n$ of state vectors over a range of operating conditions. More precisely, the fundamental assumption underlying the dimension reduction is that the $n$-dimensional state vectors $y \in \mathcal{X}$ may be approximated up to sufficient accuracy with only $p$ degrees of freedom via

$$
(8) \quad y \approx \tilde{y}(\alpha) = U_0 \alpha, \quad \alpha \in \mathbb{R}^p,
$$

where $U_0 \in St(n, p)$ is a matrix representative of $\mathcal{U}_0$. The standard case in model reduction is that the set of state vectors $\mathcal{X}$ is the solution manifold of a parametric partial differential equation (PDE).

In the following, we consider the special case that only incomplete information on a state vector $y \in \mathcal{X}$ is available. This case is encountered in the model reduction techniques gappy POD [27] and DEIM [22]. The incomplete data imposes equality

---

1For complete data vectors $b_s \in \mathbb{R}^m$, (6) is the same as [49, eq. (2)].
constraints on the $m < n$ components $y_j, \ldots, y_m$ of a state vector $y \in \mathcal{X}$ via the equation

$$P^T y \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = b, \quad P = (e_j, \ldots, e_{j_m}) \in \{0,1\}^{n \times m}.$$  

Under the requirement that $y$ be contained in $U_0$, the underdetermined equation (9) translates into the overdetermined masked least-squares problem (3) with corresponding solution (4). This establishes a direct link to the GROUSE approach.

The objective of our work is to find a subspace $U^* \in Gr(n,p)$ close to $U_0$ such that the best subspace-restricted least-squares solution $y(U^*)$ features an exact zero residual, $\|r(U^*)\|_2 = 0$. In solving this equation for the unknown $U^*$, we adapt the original reduced subspace $U_0$ according to the least-squares problem arising from the new (partial) information about $y$. The requirement of $U^*$ being close to $U_0$ is important in the context of model reduction because we want the approximation (8) to remain valid for $U^*$.

We formalize the objective. Define the feasibility set

$$\mathcal{Z} := \{U \in Gr(n,p) | \min_{y \in \mathcal{U}} \|P^T \tilde{y} - b\|_2 = 0\}. \quad (10)$$

The set $\mathcal{Z}$ is nonempty. From GROUSE, it is known that the geodesic curve $t \mapsto U(t)$ that starts in $U(0) = U_0$ with velocity given by the direction of steepest descent of the residual norm function (5) is a matrix curve of rank-one updates on the initial subspace $U_0$; see (7). We will show that this curve crosses the feasibility set $\mathcal{Z}$ and determine the first intersection point. By writing the residual vector as $r(U_0) = b - \Pi_{P^TU_0} b$, where $\Pi_{P^TU_0}$ is the orthogonal projection (2) onto colspan($P^TU_0$), this objective becomes a nonlinear equation on the Grassmann manifold:

$$b - \Pi_{P^TU(t)} b = 0 \quad (11)$$

The condition $b - \Pi_{P^TU(t)} b = 0$ is equivalent to $[U(t^*)] \in \mathcal{Z}$.

A contribution of this paper is an explicit formula for the time-dependent residual $r(U(t)) = b - \Pi_{P^TU(t)} b$ derived in section 3, from which the solution to (11) can be read off in closed form. In contrast to GROUSE, whose overall aim is the iterative global minimization of (6), we focus on the single adaptation steps and the nonlinear residual equation on $Gr(n,p)$. We arrive in this way at the same formula for $t^*$ that was obtained in [49, Alg. 1, section 3.1, App. C] as the optimal greedy step size in an iterative subspace updating scheme based on complete right-hand side vectors.

In summary, our approach is a method for determining a subspace $U^*$ contained in the set $\mathcal{Z}$ from (10) that can be reached via a geodesic path along the descent direction starting in $U_0$. Figure 1 and section SM1 from the supplement illustrate this principle. In subsection 3.3, we show that this is not restricted to the special case of masked least-squares problems $\|P^T \tilde{y} - b\|_2$ but can be generalized to arbitrary underdetermined systems $\|A\tilde{y} - b\|_2$, $A \in \mathbb{R}^{m \times n}$.

2.3. Numerical aspects of the Grassmann manifold. Our approach to solve (11) is presented in section 3 and builds on geometric concepts on the Grassmann manifold $Gr(n,p)$. This subsection reviews a few essential aspects of the numerical treatment of Grassmann manifolds. We refer the reader to [1, 2, 25] for details.

---

2 Any subspace $U$ that contains a vector $y = Pb + v$, where $v \in \mathbb{R}^n$ is in the $(n-m)$-dimensional kernel of $P^T$ is in $\mathcal{Z}$. 

---
Fig. 1. Graphical illustration of the geometric subspace adaptation: The sphere visualizes the Grassmann manifold \( \text{Gr}(n,p) \). The solid line marks the set \( Z \) of all subspaces in \( \text{Gr}(n,p) \) that contain zero-residual solutions to the least-squares problem (3). The black triangle shows the initial subspace \( U_0 \). The dashed line is the geodesic starting in \( U_0 \) with velocity given by minus the gradient of the least-squares residual function. Our goal is to compute the subspace \( U^* \), where the geodesic meets the set \( Z \).

Tangent spaces and normal coordinates. The tangent space \( T_{U_0}\text{Gr}(n,p) \) at a point \( U \in \text{Gr}(n,p) \) can be thought of as the space of velocity vectors of differentiable curves on \( \text{Gr}(n,p) \) passing through \( U \). For any matrix representative \( U \in \text{St}(n,p) \) of \( U \in \text{Gr}(n,p) \) the tangent space of \( \text{Gr}(n,p) \) at \( U \) is represented by

\[
T_{U_0}\text{Gr}(n,p) = \{ \Delta \in \mathbb{R}^{n \times p} | \ U^T\Delta = 0 \} \subset \mathbb{R}^{n \times p},
\]

its canonical metric being \( \langle \Delta, \hat{\Delta} \rangle_{\text{Gr}} = tr(\Delta^T \hat{\Delta}) \) [25, section 2.5]. Endowing each tangent space with this metric turns \( \text{Gr}(n,p) \) into a Riemannian manifold. A geodesic \( t \mapsto U(t) \) on \( \text{Gr}(n,p) \) is a locally length-minimizing curve. A geodesic is uniquely determined by its starting point \( U(0) = U_0 \in \text{Gr}(n,p) \), [2, p. 102].

The corresponding Riemannian exponential mapping is

\[
\text{Exp}_{U_0} : T_{U_0}\text{Gr}(n,p) \to \text{Gr}(n,p), \quad \Delta \mapsto \text{Exp}_{U_0}(\Delta) := U(1).
\]

The Riemannian exponential maps a tangent vector \( \Delta \in T_{U_0}\text{Gr}(n,p) \) to the endpoint \( U(1) \) of a geodesic path \( U : [0,1] \to \text{Gr}(n,p) \) starting at \( U(0) = U_0 \in \text{Gr}(n,p) \) with velocity \( \Delta \in T_{U_0}\text{Gr}(n,p) \).

An efficient algorithm for evaluating the Grassmann exponential is derived in [25, section 2.5.1]. The explicit form of the associated geodesic is

\[
U(t) = \text{Exp}_{U_0}(t\Delta) = [U_0V \cos(t\Sigma)V^T + \Phi \sin(t\Sigma)V^T], \quad \Delta \overset{\text{SVD}}{=} \Phi \Sigma V^T.
\]

The exponential mapping gives a local parametrization from the (flat, Euclidean) tangent space to the manifold. This is also referred to as representing the manifold in normal coordinates [32, section III.8], [33, Lem. 5.10].

Distance between subspaces. Given two subspaces \([U], [\hat{U}] \in \text{Gr}(n,p)\), the \( i \)th canonical or principal angle between \([U]\) and \([\hat{U}]\) is \( \theta_i := \arccos(\sigma_i) \in [0, \frac{\pi}{2}] \), where \( \sigma_i \) is the \( i \)th-largest singular value of \( U^T \hat{U} \in \mathbb{R}^{p \times p} \) [29, section 12.4.3].
The Riemannian distance between \([U], [\tilde{U}] \in Gr(n, p)\) is

\[
\text{dist}([U], [\tilde{U}]) := \|\Theta\|_2, \quad \Theta = (\theta_1, \ldots, \theta_p) \in \mathbb{R}^p.
\]

Normal coordinates are radially isometric with respect to the Riemannian distance on \(Gr(n, p)\) and the canonical metric on \(T_{I_0}Gr(n, p)\) in the following sense: the length of a tangent vector \(\Delta\) as measured by the metric in \(T_{I_0}Gr(n, p)\) is the same as the Riemannian distance \(\text{dist}(I_0, \text{Exp}_{I_0}(\Delta))\) on \(Gr(n, p)\), provided that \(\Delta\) is in a neighborhood of \(0 \in T_{I_0}Gr(n, p)\), where the exponential is invertible [33, Lem. 5.10 and Cor. 6.11].

The Grassmann manifold is a compact homogeneous space [32]. In particular, by [47, Thm. 8(b)], any two points on \(Gr(n, p)\) can be connected by a geodesic of length \(\leq \sqrt{2}\pi\). This is related to the so-called injectivity radius of the Grassmann manifold [47], which is the maximal radius \(\rho\) such that the exponential map at any point \([U] \in Gr(n, p)\) is a diffeomorphism onto the open ball \(B(0, \rho) \subset T_{[U]}Gr(n, p)\) around the origin in the corresponding tangent space. The injectivity radius of the Grassmann manifold is \(\rho = \frac{\pi}{2}\) [47]. This concept is relevant to the step of conducting the line search within Grassmann optimization schemes. We make the following observation:

Using the explicit formulas for the exponential mapping and its (local) inverse, called the logarithmic mapping \(\text{Log}_{[U]}\) (see [12, section 3]), one can show that \(\text{Log}_{[U]} \circ \text{Exp}_{[U]}(\Delta) = \Delta\) for all tangent vectors \(\Delta\) of spectral norm \(\|\Delta\|_2 = \sigma_1(\Delta) < \pi/2\), where \(\sigma_1(\Delta)\) is the largest singular value of \(\Delta\). As a consequence, we have the following observation.

**Observation 1.** For all \([U] \in Gr(n, p)\), let

\[
B_{[U], \text{spec}}(0, \pi/2) := \left\{ \Delta \in T_{[U]}Gr(n, p) \mid \sigma_1(\Delta) < \frac{\pi}{2} \right\}.
\]

Then the exponential mapping \(\text{Exp}_{[U]}\) is a radial isometry on \(B_{[U], \text{spec}}(0, \pi/2)\).

This observation is important for numerical computations because

\[
B_{[U], \text{spec}}(0, \pi/2) \supset \left\{ \Delta \in T_{[U]}Gr(n, p) \mid \sqrt{(\Delta, \Delta)_{Gr}} = \|\sigma_1, \ldots, \sigma_p\|_2^T < \frac{\pi}{2} \right\},
\]

i.e., the spectral \(\pi/2\)-ball in the tangent space encloses the canonical \(\pi/2\)-ball in the tangent space. The above observation leads to the next proposition which has implications on the uniqueness of solutions to (11).

**Proposition 1.** Let \([U] \in Gr(n, p)\), \(\Delta \in T_{[U]}Gr(n, p)\), and \(\tilde{U} = \text{Exp}_{[U]}(\Delta)\). If \(\|\Delta\|_2 < \frac{\pi}{2}\), then \(\text{dist}([U], [\tilde{U}]) = \|\Delta\|_{Gr}\). In particular, the length of the geodesic path starting in \([U]\) and ending in \([\tilde{U}]\) is less than \(\sqrt{2}\pi\).

**Proof.** Let \(\Delta \overset{\text{SVD}}{=} \Phi \Sigma V^T\) with \(\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_p)\) and let \(\sigma_1 = \|\Delta\|_2 < \frac{\pi}{2}\). The exponential projection of \(\Delta\) onto \(Gr(n, p)\) is \([\tilde{U}] = \text{Exp}_{[U]}(\Delta) = [UV \cos(\Sigma)V^T + \Phi \sin(\Sigma)V^T]\).

The SVD of \(U^T \tilde{U}\) is \(V \cos(\Sigma)V^T\), so that \(0 \leq \theta_k := \arccos(\cos(\sigma_k)) = \sigma_k < \frac{\pi}{2}\). Hence, \((\sigma_1, \ldots, \sigma_p)^T = (\theta_1, \ldots, \theta_p)^T := \Theta \in \mathbb{R}^p\) is precisely the vector of canonical angles between \([U]\) and \([\tilde{U}]\) (when listing the canonical angles in descending order); see (13). As a consequence,

\[
\text{dist}([U], [\tilde{U}]) = \|\Theta\|_2 = \sqrt{\text{tr}(\Sigma^2)} = \sqrt{\text{tr}(\Delta^T \Delta)} = \|\Delta\|_{Gr}.
\]

Since \(\sigma_1 < \frac{\pi}{2}\), we have \(\|\Delta\|_{Gr} = \left(\sum_{i=1}^p \sigma_i^2\right)^{1/2} < \frac{\sqrt{2}\pi}{2}\).
A subtlety of Proposition 1 is that the length condition on $\Delta$ is with respect to the spectral norm rather than the canonical norm.

3. Solving the Grassmann residual equation. We now return to our goal formulated in subsection 2.2: the solution of (11). In subsection 3.1, we derive a general update formula for orthogonal projectors under rank-one modifications. Subsection 3.2 derives an explicit time-dependent expression for the Grassmann residual along the GROUSE geodesic. In particular, this allows us to read off the closed-form solution to (11). A generalization to least-squares systems featuring arbitrary matrices rather than mask matrices as operators is given in subsection 3.3. Subsequently, subsection 3.4 introduces an extension for performing the Grassmann subspace adaptation over selected directions of the subspace only.

3.1. A closed-form rank-one update for orthogonal projectors. In this subsection, we derive a formula for orthogonal projectors under rank-one updates that turns out to be an essential building block in solving (11). As this result is also of independent interest, we state it in a more general setting.

Let $X \in \mathbb{R}^{m \times p}$. Recall from (2) that the orthogonal projection onto $\text{colspan} \ X$ is $\Pi_X = X(X^TX)^{-1}X^T$. Let $X^{SVD} = Q\Sigma R^T$ be the thin SVD of $X$ with $Q \in S \{m, p\}$, $\Sigma \in \mathbb{R}^{p \times p}$ diagonal, $R \in \mathcal{O}_p$ orthogonal. Then $\Pi_X$ is expressed alternatively as $\Pi_X = QQ^T$.

Let $x \in \mathbb{R}^m$, $v \in \mathbb{R}^p$, and consider the rank-one update $X_{new} = X + xv^T \in \mathbb{R}^{m \times p}$. We are interested in an expression $\Pi_{X_{new}} = Q_{new}^{T}Q_{new}$, where $Q_{new} \in S \{m, p\}$. One standard way to approach this is via rank-one SVD updates [19, 17]. However, this requires an auxiliary SVD of a $(p \times p)$-matrix. Here, we can avoid this, since we are not interested in the fully updated $X_{new}^{SVD} = Q_{new}\Sigma_{new}R_{new}^{T}$ or even in $Q_{new}$ alone but only in $Q_{new}^{T}$.

**Lemma 2.** As in the above setting, let $X \overset{SVD}{=} Q\Sigma R^T$, $X_{new} = X + xv^T$, and define

$$\begin{align*}
\hat{q} &= x - QQ^T x, \quad q = \frac{\hat{q}}{\|\hat{q}\|_2} \in \mathbb{R}^m, \\
g &= \begin{pmatrix} g_p \\ g_{p+1} \end{pmatrix} = \begin{pmatrix} -\Sigma^{-1}R^Tv \\ \frac{1}{\|q\|_2} \left(1 + x^TQ\Sigma^{-1}R^Tv\right) \end{pmatrix} \in \mathbb{R}^{p+1}.
\end{align*}$$

Then the orthogonal projection onto $\text{colspan}(X_{new})$ is

$$\Pi_{X_{new}} = (Q, q) \begin{pmatrix} Q^T \\ q^T \end{pmatrix} - \frac{1}{\|g\|_2^2} (Q, q)gg^T \begin{pmatrix} Q^T \\ q^T \end{pmatrix}.$$

**Proof.** We start with a decomposition inspired by [17, eq. (3)]. Note that $(Q, q) \in S \{m, p+1\}$ by construction. It holds that

$$X + xv^T = (Q, q) \begin{pmatrix} \Sigma R^T + Q^T xv^T \\ \hat{q}v^T \end{pmatrix} =: (Q, q)M,$$

where $M \in \mathbb{R}^{(p+1) \times p}$. Let $M^{SVD} = \tilde{Q}\tilde{\Sigma}\tilde{R}^T$ be the thin SVD of $M$, i.e., $\tilde{Q} \in S \{p+1, p\}$, $\tilde{\Sigma}, \tilde{R}^T \in \mathbb{R}^{p \times p}$. Formally, the updated SVD is

$$X + xv^T = \left((Q, q)\tilde{Q}\right)\tilde{\Sigma}\tilde{R}^T =: Q_{new}\Sigma_{new}R_{new}^{T}.$$
Let $g \in \mathbb{R}^{p+1}$ be such that $(\tilde{Q}, \frac{g}{\|g\|}) \in O_{p+1}$ is an orthogonal completion of $\tilde{Q}$. Because of $I_{p+1} = (\tilde{Q}, \frac{g}{\|g\|})(\tilde{Q}, \frac{g}{\|g\|})^T$, we have

$$\hat{Q}\hat{Q}^T = I_{p+1} - \frac{1}{\|g\|^2}gg^T$$

and, as a consequence,

$$(16) \quad Q_{new}^TQ_{new} = (Q, q)\hat{Q}\hat{Q}^T \left(\frac{Q^T}{q^T}\right) = (Q, q) \left(I_{p+1} - \frac{1}{\|g\|^2}gg^T\right) \left(\frac{Q^T}{q^T}\right).$$

Hence, it is sufficient to determine $g$, which is characterized up to a scalar factor by $\hat{Q}^T g = 0$. Since $\text{colspan}(M) = \text{colspan}(\hat{Q})$, this condition is equivalent to $M^T g = 0$. Let $g_p \in \mathbb{R}^p$ denote the first $p$ components of $g$ and let $g_{p+1} \in \mathbb{R}$ be the last entry such that $g^T = (g_p^T, g_{p+1})$. When writing the equation $g^T M = 0$ as

$$(g_p^T, g_{p+1}) \left(\begin{array}{c} \Sigma \\ 0 \\ \|q\|_2 \end{array}\right) \left(\begin{array}{c} R^T \\ v^T \end{array}\right) = 0,$$

it is straightforward to show that

$$g = \left(\begin{array}{c} \frac{-\Sigma^{-1} R^Tv}{\|q\|_2} \\ \frac{1}{\|q\|_2} (1 + x^T \Sigma^{-1} R^Tv) \end{array}\right) \in \mathbb{R}^{p+1}$$

and any scalar multiple of this vector is a valid solution. Using this vector in (16) proves the lemma. \qed

3.2. An explicit expression for the Grassmann residual function along the GROUSE geodesic. We now state our main theorem on the solution of the nonlinear equation (11).

**Theorem 3.** Let $U_0 = [U_0] \in \text{Gr}(n, p)$ be represented by $U_0 \in \text{St}(n, p)$. Let $P = (e_{j_1}, \ldots, e_{j_m}) \in \{0, 1\}^{(n \times m)}$ be a mask matrix. Moreover, let $b \in \mathbb{R}^m$ and suppose that $U_0^T Pb \neq 0$.

Let $\alpha = (U_0^TPP^TU_0)^{-1}U_0^T Pb$ be the optimal coefficient vector corresponding to the masked least-squares problem

$$\min_{\alpha \in \mathbb{R}^p} \|P^TU_0\alpha - b\|^2$$

and let $r = b - P^TU_0\alpha$ the associated residual vector. Set $v = \frac{\alpha}{\|\alpha\|_2}$ and $s_1 = 2\|r\|_2\|\alpha\|_2$. Moreover, write $P^TU_0 \overset{\text{SVD}}{=} Q\Sigma R^T \in \mathbb{R}^{n \times p}$ and $g_p = -\Sigma^{-1} R^Tv$. The $t$-dependent residual vector along the geodesic descent direction is

$$r([U(t)]) = b - \Pi_{P^TU(t)} b = \|r\|_2 - \|\alpha\|_2 \tan(s_1) \frac{r}{\|r\|_2} - \frac{\tan(s_1)}{\|\alpha\|_2} Q\Sigma^{-2} Q^T b.$$ 

**Proof.** Reconsider (7) and let

$$x(t) = P^T \dot{x}(t) = (\cos(ts_1) - 1)P^TU_0 v + \sin(ts_1) \frac{r}{\|r\|_2},$$

$$v = \frac{\alpha}{\|\alpha\|_2}, \quad \alpha = (U_0^TPP^TU_0)^{-1}U_0^T Pb,$$
so that

\[ P^T U(t) = P^T U_0 + x(t)v^T. \]

Since \( P^T U(t) \) is a rank-one update of \( P^T U_0 \), Lemma 2 applies. Introducing \( P^T U_0 = \text{SVD} Q\Sigma R^T \in \mathbb{R}^{m \times p} \), we obtain \( r = b - QQ^T b \) and \( \alpha = R\Sigma^{-1}Q^T b \). The \( t \)-dependent orthogonal projection onto \( \text{colspan}(P^T U(t)) \) is

\[
\Pi_{P^T U(t)} = (Q, q(t)) \left( \frac{Q^T}{q^T(t)} \right) - \frac{1}{\|g(t)\|_2^2} (Q, q(t))g(t)g^T(t) \left( \frac{Q^T}{q^T(t)} \right),
\]

where

\[
\tilde{q}(t) = x(t) - QQ^T x(t), \quad q(t) = \frac{\tilde{q}(t)}{\|\tilde{q}(t)\|_2} \in \mathbb{R}^m,
\]

\[
g(t) = \left( \frac{g_p}{g_{p+1}(t)} \right) = \left( \frac{1}{\|\tilde{q}(t)\|_2} \frac{-\Sigma^{-1}R^Tv}{(1 + x^T(t)\Sigma^{-1}R^Tv)} \right) \in \mathbb{R}^{p+1}.
\]

We have \( Q^T r = 0 \) and thus \( Q^T x(t) = \frac{\cos(ts_1) - 1}{\|\alpha\|_2^2} Q^T b \). This leads to \( \tilde{q}(t) = \frac{\sin(t)}{\|\alpha\|_2} r \) and \( \|\tilde{q}(t)\|_2 = |\sin(t)| \) as well as \( q(t) = \text{sign}(\sin(t)) \frac{r}{\|\alpha\|_2} = \pm q \), where we standardize \( q = \frac{r}{\|\alpha\|_2} \). Moreover,

\[
x^T(t)Q\Sigma^{-1}R^Tv = \frac{1}{\|\alpha\|_2^2} (\cos(ts_1) - 1) b^T Q\Sigma^{-2}Q^T b = (\cos(ts_1) - 1),
\]

so that \( g(t) \) is

\[
g(t) = \left( -\frac{1}{\|\alpha\|_2^2} \frac{\Sigma^{-2}Q^T b}{\cos(ts_1)} \right) \in \mathbb{R}^{p+1}.
\]

It holds that \( \frac{\cos(ts_1)}{|\sin(ts_1)|} q(t) = \frac{\cos(ts_1)}{|\sin(ts_1)|} g(t) \). Hence, according to (18), we may consistently work with \( \pm q \) and \( \frac{\cos(ts_1)}{|\sin(ts_1)|} = \cot(ts_1) \). In order to evaluate the updated projection (18), we compute

\[
(Q, q) \left( \frac{g_p}{g_{p+1}(t)} \right) = -\frac{1}{\|\alpha\|_2^2} Q\Sigma^{-2}Q^T b + \cot(ts_1)q,
\]

\[
g_p^T Q^T b = -\frac{1}{\|\alpha\|_2^2} b^T Q\Sigma^{-2}Q^T b = -\|\alpha\|_2, \text{ and}
\]

\[
q^T b = \frac{1}{\|r\|_2^2} r^T b = \frac{1}{\|r\|_2^2} \left( \frac{b^T b - b^T Q^T b}{\|r\|_2^2} \right) = \|r\|_2.
\]

Substituting these identities in (18), we arrive at

\[
r([U(t)]) = b - \Pi_{P^T U(t)} b = b - QQ^T b - qq^T b
\]

\[
+ \frac{1}{\|g(t)\|_2^2} (Qg_p + \cot(ts_1)q) \left( g_p^T Q^T b + \cot(ts_1)q^T b \right)
\]

\[
= \frac{\cot(ts_1)}{\|r\|_2^2 - \|\alpha\|_2^2} \left( \cot(ts_1) \frac{r}{\|r\|_2} - \frac{1}{\|\alpha\|_2^2} Q\Sigma^{-2}Q^T b \right),
\]

as was claimed.
Note that the only special property of $P$ that is exploited in the proof is that $P^TPr = r$. Hence, the result holds when $P$ is replaced with an arbitrary column-orthogonal matrix.

There is a number of conclusions that can be drawn from Theorem 3.

**Corollary 4.**

1. The $t$-dependent residual norm along the steepest descent direction is

$$
||r([U(t)])||_2 = \left|\begin{array}{cc}
\|b - \Pi_{P^TU(t)}b\|_2 &= \frac{||r||_2 - \|\alpha\|_2 \tan(ts_1)}{\sqrt{1 + \|g_p\|^2 \tan^2(ts_1)}}.
\end{array}\right.
$$

2. The residual norm function is continuous and $\frac{\pi}{s_1}$-periodic along the steepest descent direction with

$$
||r([U(0)])||_2 = ||r||_2 \text{ and } ||r\left([U\left(\frac{\pi}{2s_1}\right)\right])||_2 = \frac{\|\alpha\|_2}{\|g_p\|_2} = \frac{\|\alpha\|_2}{\|Q\Sigma^{-2}Q^Tb\|_2}.
$$

3. The first root along the geodesic descent direction is at

$$
t^* = \frac{1}{s_1}\text{arctan}\left(\frac{||r||_2}{\|\alpha\|_2}\right) \in \left(0, \frac{\pi}{2s_1}\right).
$$

The associated matrix $U^* := U_0 + ((\cos(ts_1) - 1)U_0v + \sin(ts_1)\frac{F_v}{\|v\|^2})v$ is such that the subspace $U^* := [U^*]$ is contained in the set $Z$ from (10), i.e.,

$$
F(U^*) = \min_{\alpha \in \mathbb{R}^n} \|P^TU^*\alpha - b\|^2 = 0.
$$

Stated differently, it holds that $b$ is contained in $\text{colspan}(P^TU^*)$, that is, $b = \Pi_{P^TU^*}b$.

4. The coefficient vector associated with $U^* = [U^*] = (23)$ is $\alpha^* = \sqrt{1 + \frac{\|r\|^2}{\|\alpha\|^2}}\alpha$.

The associated $y^* \in \mathbb{R}^n$ is $y^* = U^*\alpha^* = U_0\alpha + Pr = U_0\alpha + P(b - P^TU_0\alpha)$. Hence, $y^*$ can be readily obtained without computing any of $t^*, \alpha^*, U^*$.

5. The first maximum along the geodesic descent direction is at

$$
t_{\text{max}} = \frac{1}{s_1}\left(\pi - \text{arctan}\left(\frac{\|\alpha\|_2}{\|r\|^2\|g_p\|^2}\right)\right) \in \left(\frac{\pi}{2s_1}, \frac{\pi}{s_1}\right)
$$

with corresponding value $||r([U(t_{\text{max}})])||_2 = \sqrt{\|r\|^2 + \frac{\|\alpha\|^2}{\|g_p\|^2}}$.

**Proof.** By taking into account that $r$ is orthogonal to $\text{colspan}(Q)$, Pythagoras' theorem gives $\|((\cot(ts_1)\frac{r}{\|r\|^2} - \frac{1}{\|\alpha\|^2}Q\Sigma^{-2}Q^Tb)\|_2 = \sqrt{\cot^2(ts_1) + \|g_p\|^2} = \|g(t)\|_2$. The formula (21) is now an immediate consequence of (20). From (21), statements 2, 3, and 5 of the corollary are straightforward.

On statement 4: From statement 3, we know that there exists $\alpha^* \in \mathbb{R}^p$ such that $P^TU^*\alpha^* - b = 0$. After plugging in the explicit expression for $U^*$, we obtain the equation

$$
P^TU_0\left(\alpha^* - \frac{\alpha^T\alpha^*}{\|\alpha\|^2}\alpha\right) + \left(\frac{\alpha^T\alpha^*}{\|\alpha\|^2\sqrt{\|\alpha\|^2 + \|r\|^2}} - 1\right)b = 0.
$$

If the unmodified least-squares problem (3) features a nonzero residual, then $b$ is not contained in $\text{colspan}(P^TU_0)$. Hence, both quantities in the round brackets must be zero, which leads to $\alpha^* = \frac{\alpha^T\alpha^*}{\|\alpha\|^2}\alpha = \frac{\sqrt{\|\alpha\|^2 + \|r\|^2}}{\|\alpha\|^2}\alpha$. The calculation of $y^*$ is, therefore, straightforward. \qed
Appendix A features a shortcut to statements 3 and 4 of Corollary 4. An example of a plot of the residual norm function (21) from a practical application is displayed in Figure 5.

Remark 5. The GROUSE convergence analysis in [10] is based on local considerations and a step length of \( t = \frac{1}{s_1} \arcsin \left( \frac{|r|}{\|\alpha\|} \right) \), which matches the \( t^* \) in (22) up to terms of third order, when the residual and, therefore, the ratio \( \|r\|/\|\alpha\| \) are small. In the fully sampled case, that is, when complete right-hand side data is available, [49] shows that the same \( t^* \) of (22) is also the greedy-optimal step with respect to the determinant-similarity and the Frobenius norm discrepancy of two subspaces in an iterative subspace updating scheme; see [49, section 3.1 and App. C]. In contrast, we arrived at \( t^* \) from the independent approach of solving the nonlinear equation (11) and with a different proof that relies on Lemma 2. Combining these facts shows that the subspace discrepancy is maximal if and only if the subspace update is such that the residual vanishes exactly.

The proof of Proposition 1 shows that the distance between the subspaces \([U_0]\) and \([U^*]\) is \( t^* s_1 = \arcsin \left( \frac{|r|}{\|\alpha\|} \right) < \frac{\pi}{2} \). Hence, when performing the \( t^* \)-optimal rank-one update on \([U_0]\) according to Corollary 4, we stay within the injectivity radius. As a consequence from general differential geometry, the geodesic \( t \mapsto [U(t)] \) is length-minimizing, that is, there is no shorter curve on \( Gr(n, p) \) that connects \([U_0]\) and \([U^*]\).

We emphasize that the update formula of Lemma 2 for orthogonal projectors under rank-one modifications was used as an intermediate theoretical fact in proving Theorem 3 but that it is not required to actually compute the rank-one update and the associated quantities \( Q, q, g \) in order to obtain the optimal \( t^* \) and the subspace \([U^*] = [U(t^*)]\). MATLAB code that considers this fact is in the supplement section SM4.

We draw a corollary that corresponds to the special case where the mask matrix \( P \) is the identity \( I_n \), i.e., the case where complete data is available. Recall that the best least-squares approximation to a given vector \( b \) that is contained in a subspace \( \mathcal{U}_0 \) is the orthogonal projection \( U_0 U_0^T b \) of \( b \) onto \( \mathcal{U}_0 \), with an associated residual of \( r = b - U_0 U_0^T b \). The SVD of \( P^T U_0 \) is now trivially \( P^T U_0 = Q \Sigma R^T = U_0 I_p I_p^T \) so that the expressions involving \( Q, \Sigma, R \) simplify.

Corollary 6. Let \( \mathcal{U}_0 = [U_0] \in Gr(n, p) \) be represented by \( U_0 \in St(n, p) \). Let \( b \in \mathbb{R}^n \) and suppose that \( \alpha := U_0^T b \neq 0 \). Set \( v = \frac{a}{\|\alpha\|} \) and \( s_1 = 2 \|r\|/\|\alpha\| \). Then the \( t \)-dependent residual norm is

\[
\|r([U(t)])\|_2 = \|b - \Pi_{U(t)} b\|_2 = \frac{\|r\|_2 - \|\alpha\|_2 \tan(ts_1)}{\sqrt{1 + \tan^2(ts_1)}}.
\]

Define

\[
t^* = \frac{1}{s_1} \arctan \left( \frac{\|r\|_2}{\|\alpha\|_2} \right).
\]

Then \( U^* := U(t^*) := U_0 + ((\cos(t^* s_1) - 1)U_0 v + \sin(t^* s_1) \frac{r}{\|r\|}) v^T \) is such that \( b \) is contained in the subspace \( U^* := [U^*] \), i.e., \( b = \Pi_{U^*} b \).

Remark 7. Corollary 6 has a connection with rank-one SVD updates as considered in [19, 16, 17]. One application in [17, Table 1] is to revise an existing SVD

---

This does not necessarily mean that there is no other point \([U^*] \in \mathcal{Z} \) that is closer to \([U_0]\).
squares problem \( Q \) with \( Q \) restricted least-squares problems \( P \) is replaced with an arbitrary column-orthogonal matrix. For general subspace-

tinuous to apply. Convergence results for GROUSE with arbitrary sampling matrices 

\[
\text{gradient associated with the residual function is still rank-one so that GROUSE con-}
\]

\[
\text{problem (3) is not a mask matrix but an arbitrary real matrix, then the Grassmann}
\]

\[
\text{supplemental section SM2.}
\]

\[
\text{dimensions. It is shown that this procedure can be emulated via GROUSE when}
\]

\[
\text{the augmented matrix representative and consequently retruncated to its original}
\]

\[
\text{data to a given subspace representative. Then, the SVD update is performed on}
\]

\[
\text{in [9]. The approach considered in [9] corresponds to first attaching new column}
\]

\[
\text{in the update scheme enables more general updates than when working with a fixed}
\]

\[
\text{U obtained in this way does not feature the property}
\]

\[
\text{a specific step size is chosen for the rank-one increment. However, the modified}
\]

\[
\text{U in the update scheme enables more general updates than when working with a fixed}
\]

\[
\text{U represent.}
\]

\[
\text{0 = } \| Q^T U^* \tilde{\alpha} - (R^T)^{-1}b \|^2. \quad \text{As a consequence, } \| AU_0 \tilde{\alpha} - b \|^2 = 0. \quad \text{In summary, we}
\]

\[
\text{present the following theorem.}
\]

\[
\text{Theorem 8. Let } p < m \leq n. \text{ Consider the general subspace restricted least-}
\]

\[
\text{squares problem}
\]

\[
\min_{\tilde{\alpha} \in \mathbb{R}^p} \| AU_0 \tilde{\alpha} - b \|^2, \quad A \in \mathbb{R}^{m \times n}, \quad b \in \mathbb{R}^m, \quad [U_0] \in Gr(n, p), \quad \text{rank}(AU_0) = p.
\]
Let \( QR = A^T \) and suppose that \( R \) is regular. Then there exists a subspace \([U^*] \in Gr(n, p)\) such that

\[
\min_{\alpha \in \mathbb{R}^p} ||AU^*\alpha - b||^2 = 0 \quad \text{and} \quad \text{dist}([U_0], [U^*]) = \arctan\left(\frac{||r||_2}{||\alpha||_2}\right) = \tau^*,
\]

where \( \alpha = \arg \min_{\alpha \in \mathbb{R}^p} ||Q^TU_0\alpha - (R^T)^{-1}b||^2, \ r = (R^T)^{-1}b - Q^TU_0\alpha \). The subspace \( U^* \) is given by

\[
U^* = U_0 + \left((\cos(\tau^*) - 1)U_0\frac{\alpha}{||\alpha||_2} + \sin(\tau^*)\frac{Qr}{||r||_2}\right)\frac{\alpha^T}{||\alpha||_2}.
\]

### 3.4. Adapting a subspace of a subspace.

There are many applications where it might be desirable to keep some directions of a given subspace fixed while adapting the remaining ones. In the context of adaptive model reduction, such situations are likely to occur if the columns spanning the subspace in question stem from a principal component analysis or proper orthogonal decomposition (POD), and are thus ordered by information content. In these cases, the user might want to keep the most dominant subspace directions fixed, while adapting the portion of the subspace spanned by the less important basis vectors. This subsection describes the modifications to the methodology for doing so; a sample application is presented in subsection 4.2.

Let \( f : Gr(n, p) \rightarrow \mathbb{R}, [U] \mapsto f([U]) \) be a differentiable function. Let us divide the column set of a subspace representative \( U \in St(n, p) \) into a constant portion \( U_c \in St(n, p - l) \) and a portion \( U_l \in St(n, l) \) that is subject to change, so that \( U = (U_c, U_l) \in St(n, p - l) \times St(n, l) \). By fixing \( U_c \), we obtain a function \( f_1 : Gr(n, l) \rightarrow \mathbb{R}, f_1([U_l]) = f([U_c, U_l]) \) with gradient \( G_1 := \nabla f_1([U_l]) \in \mathbb{R}^{n \times l} \). The gradient induces the search direction \( H_l = -G_1 \). The geodesic associated with the search direction \( H_l \) is represented by

\[
\Phi_1 S_1 V_l^T \in \mathbb{R}^{n \times l}
\]

\[(24) \quad U_l(t) = \exp([U_l]) t H_l = U_l(t) V_l \cos(t S_l) V_l^T + \Phi_1(t) \sin(t S_l) V_l^T.
\]

Note that \( S_l \) and \( V_l \) are \((l \times l)\)-matrices. For each \( t \), the matrix \( U_l(t) \in St(n, l) \) is a feasible orthogonal subspace representative. Yet, we have to consider the possibility that the compound matrix \((U_c, U_l(t))\) ceases to be a valid subspace representative in \( St(n, p) \).\(^4\) It is even conceivable that \([U_l(t)]\) moves towards the subspace \([U_c]\) spanned by the fixed basis vectors so that the compound matrix \((U_c, U_l(t))\) no longer loses the orthogonal-columns property but even becomes rank deficient. One way to avoid this is to reorthogonalize \( U_l(t) \) against \( U_c \), say, by conducting an extra Gram–Schmidt procedure. However, Proposition 9 below implies that the orthogonality between the columns of the matrices \( U_l(t) \) and the constant columns of the matrix block \( U_c \) is preserved along the geodesic path in direction of the least-squares gradient, so that in this case, the corresponding compound matrix \((U_c, U_l(t))\) is also an orthogonal matrix representative in \( St(n, p) \) and a Gram–Schmidt reorthogonalization is unnecessary.

**Proposition 9.** Let \( f : Gr(n, p) \rightarrow \mathbb{R} \) be differentiable. Suppose that

\[(25) \quad T_{[U]}Gr(n, p) \ni \nabla_{[U]} f = (\nabla_{[U_c]} f_c, \nabla_{[U_l]} f_l) \in (T_{[U_c]}Gr(n, p - l)) \times (T_{[U_l]}Gr(n, l)),
\]

where it is understood that \( \nabla_{[U_c]} f_c \) and \( \nabla_{[U_l]} f_l \) denote the gradients of the restrictions \( f_c : [U_c] \rightarrow f([U_c, U_l]) \) and \( f_l : [U_l] \rightarrow f([U_c, U_l]) \), respectively.

Let \([U_0] = [U_c, U_{0,l}] \in Gr(n, p)\). Let \( t \mapsto [U_l(t)] \subset Gr(n, l) \) be the geodesic path along the descent direction \(-\nabla_{[U_l]} f_l\). Then \( U_l^T U_l(t) = 0 \) for all \( t \).

\(^4\) Appendix B shows that this actually may happen even along search directions of rank one.
Therefore, the corresponding curve of concatenated matrices \((U_c, U_l(t)) \subset \mathbb{R}^{n \times p}\) is a curve of orthogonal matrices in \(St(n, p)\). Hence, for each \(t\), \([(U_c, U_l(t))] \in Gr(n, p)\), consistent with the quotient space viewpoint (1).

Proof. Let \(U_0 = (U_c, U_{l, 0}) \in St(n, p)\), where \(U_c \in St(n, p - l)\) and \(U_{l, 0} \in St(n, l)\).

The gradient with respect to \(f\) is a tangent vector in \(T_{[U_0]} Gr(n, p)\), hence \(U_0^T \nabla_{[U_0]} f = 0\). By (25),

\[
0 = U_0^T \nabla_{[U_0]} f = \left( \begin{array}{c} U_c^T \\ U_l^T \\ \end{array} \right) \left( \nabla_{[U_c]} f_c, \nabla_{[U_{l, 0}]} f_l \right).
\]

\[(26)\]

In particular, \(U_l^T \nabla_{[U_{l, 0}]} f_l = 0\). Writing \(\nabla_{[U_{l, 0}]} f_l = \Sigma \Phi_l V_l^T \in \mathbb{R}^{n \times l}\), we have \(U_l^T \Phi_l = 0\), since the columns of \(\Phi_l\) span the same space as the columns of \(\nabla_{[U_l]} f_l\). Hence, the geodesic at \(t\), \(U_l(t) = U_{l, 0} V_l \cos(t S_l) V_l^T + \Phi_l \sin(t S_l) V_l^T\) is also orthogonal to \(U_c\), i.e., \(U_c^T U_l(t) = 0\).

As can be seen from the proof, the proposition is not specific to the GROUSE context nor does it depend on the rank of the gradient. It holds, in general, whenever the gradient splitting of (25) holds. This, however, is not always the case; see Appendix B. The objective function \(F\) of (5) features this property: When allowing only the last \(l\) directions of \((U_c, U_l)\) to vary, we obtain a differentiable \(F_l : Gr(n, l) \rightarrow \mathbb{R}\) with

\[
F_l([U_l]) = b^T b - b^T P T (U_c, U_l) \left( \begin{array}{c} U_c^T \\ U_l^T \\ \end{array} \right) P P^T (U_c, U_l)^{-1} \left( \begin{array}{c} U_c^T \\ U_l^T \\ \end{array} \right) P b.
\]

The associated gradient, now a rank-one \((n \times l)\)-matrix, reads

\[
G_l := \nabla_{[U_l]} F_l = -2 P \left( b - P P^T U \alpha \right) \alpha^T \left( \begin{array}{c} 0_{(p-l) \times l} \\ I_l \end{array} \right), \quad \alpha = (U_c^T P P^T U)^{-1} U_l^T P b,
\]

where \(U = (U_c, U_l)\). The next corollary transfers the result of Corollary 4 to the setting of adapting only the last \(l\) columns of a given subspace representative.

**Corollary.** Let \(U_0 = [U_0] \in Gr(n, p)\) be represented by \(U_0 \in St(n, p)\). Let \(P = \{e_{j_1}, \ldots, e_{j_m}\} \in \{0, 1\}^{(n \times m)}\) be a mask matrix and let \(b \in \mathbb{R}^m\).

Let \(\alpha = (U_0^T P P^T U_0)^{-1} U_0^T P b\) be the optimal coefficient vector corresponding to the masked least-squares problem

\[
\min_{\alpha \in \mathbb{R}^p} \|P T U_0 \tilde{\alpha} - b\|^2
\]

and let \(r = b - P T U_0 \alpha\) be the associated residual vector. Let \(l \in \mathbb{N}, l \leq p\) and write columnwise \(U_0 = (U_c, U_{l, 0}), U_c = (u_0^1, \ldots, u_0^{p-l}), U_{l, 0} = (u_0^{p-l+1}, \ldots, u_0^l)\). Moreover, let \(\alpha_l = (0_{(p-l) \times l}, I_l)\) \(\alpha\) and \(v_l = \frac{\alpha_l}{\|\alpha_l\|_2} \in \mathbb{R}^l\).

Set \(s_1 = 2r_2\|\alpha_l\|_2\) and define

\[
t^* = \frac{1}{s_1} \arctan \left( \frac{\|r\|_2}{\|\alpha_l\|_2} \right)
\]

and \(U_l(t^*) = U_{l, 0} + ((\cos(t^* s_1) - 1) U_l v_l + \sin(t^* s_1) \frac{P r}{P r^T}) v_l^T\).

Then \(U^* := U(t^*) = (U_c, U_l(t^*))\) is such that the subspace \(U^* := [U^*]\) is contained in the set \(Z\) from (10), i.e.,

\[
F([U^*]) = \min_{\alpha \in \mathbb{R}^p} \|P T U^* \tilde{\alpha} - b\|^2 = 0,
\]

which means that \(t^*\) solves (11).
Proof. According to Proposition 9, the concatenated matrix \((U_c, U_l(t))\) is a valid subspace representative in \(S(t(n, p))\) for all \(t\). Applying the mask operator \(P\) to \((U_c, U_l(t))\) leads to the matrix curve

\[
P^T U(t) = P^T (U_c, U_l(t)) = P^T (U_c, U_l V_l \cos(tS_l)V_l^T + \Phi_l \sin(tS_l)V_l^T).
\]

Because \(\Phi_l, S_l, V_l\) stem from an SVD of the rank-one gradient \(-G_l\), we have that \(S_l = \text{diag}(s_1, 0, \ldots, 0), s_1 = 2\|r\|_2\|\alpha_l\|_2\). It follows that

\[
P^T U(t) = P^T (U_c, U_l, 0) + \left(0_{n \times (p-l)}, (\cos(ts_l) - 1)P^T U_l v_l + \sin(ts_l)\frac{r}{\|r\|_2} \right)
\]

\[
= P^T (U_c, U_l, 0) + x(t) \left(0_{1 \times (p-l)}, v_l^T \right),
\]

where \(v_l = \alpha_l\|\alpha_l\|_2^{-1}\) is the first column of \(V_l\). This is again a rank-one update on \(P^T U(t)\) and the rest of the proof is analogous to the proof of Theorem 3. \(\square\)

Remark 11. When we are adapting only the last column \(u_0^l\) of the initial matrix \(U_0 = (u_0^1, \ldots, u_0^n) \in S(t(n, p))\), then the resulting \(U^*\) is given by \((u_0^1, \ldots, u_0^{p-1}, u_0^p(t^*))\), where the last column evaluates to \(u_0^p(t^*)^{*} = \frac{1}{\|r\|_2 + \|\alpha_l\|_2}(u_0^p\alpha_p + Pr)\). This is precisely the same result that is obtained by replacing the last column of \(U_0\) with the vector \(U_0\alpha + Pr\) and reorthogonalizing the new last column against the columns of \(U_0^{p-1} := (u_0^1, \ldots, u_0^{p-1})\) via a single Gram–Schmidt step \((I - U_0^{p-1}(U_0^{p-1})^T)(U_0\alpha + Pr) = (u_0^p\alpha_p + Pr)\). In this case, the \(t^*\)-GROUSE update applied to the last column of the subspace representative \(U_0\) and the \([U]\)-part of the “revise” SVD update of \([17, Table 1, p. 23]\) coincide; cf. Remark 7. For more details, see supplemental section SM2.

4. Application to adaptive model reduction. This section applies the geometric rank-one subspace update in the specific contexts of online adaptive model reduction and image reconstruction. For each application, we describe how the subspace adaptation is employed and we demonstrate the method with numerical examples.

4.1. Adaptation for POD-DEIM reduced models. We present an online adaptive DEIM that is based on our geometric rank-one subspace update. In contrast to the standard use case in the GROUSE literature \([8, 49]\), the focus here is not on estimating a subspace from scratch based on a global objective function \((6)\) but adapting a subspace that is already a good approximant for the underlying simulation process during the online phase.

We first formulate our online adaptive DEIM for nonlinear dynamical systems and then present numerical results for the FitzHugh–Nagumo system. To ease exposition and to focus on benchmarking our online adaptive DEIM reduced models, we consider dynamical systems without parameters and inputs. Thus, the aim of the following reduced models is to reproduce well the solution of the full-order dynamical system, instead of predicting solutions for new parameters and inputs. We note, however, that the following POD-DEIM and our online adaptive POD-DEIM reduced models are applicable to parametrized models and models with inputs; see \([43, 14]\).

4.1.1. POD-DEIM-Galerkin reduced models. Consider a nonlinear dynamical system in the time interval \([0, T] \subset \mathbb{R}\), with end time \(T > 0\). Let \(t_0, t_1, \ldots, t_K \in [0, T] \subset \mathbb{R}\) be \(K + 1 \in \mathbb{N}\) time steps with \(t_0 = 0\) and \(t_K = T\). Discretizing with, e.g., the forward Euler method leads to the system of equations

\[
Ey_i = Ay_{i-1} + f(y_{i-1}), \quad i = 1, \ldots, K,
\]
corresponding to the time steps $t_1, \ldots, t_K$, respectively. Let $n \in \mathbb{N}$ denote the dimension of the discrete state space. We have the system matrices $A \in \mathbb{R}^{n \times n}$ and $E \in \mathbb{R}^{n \times n}$. The nonlinear function $f : \mathbb{R}^n \to \mathbb{R}^n$ corresponds to the nonlinear terms of the dynamical system. The state vector at time step $t_i$ is denoted as $y_i \in \mathbb{R}^n$. The initial condition is $y_0 \in \mathbb{R}^n$. We consider here the case where the nonlinear function $f$ is evaluated componentwise at the state vector $y_i$; see, e.g., [22]. We further assume the well-posedness of (27).

To derive a reduced model of the full model (27), we select a set of $n_s \in \mathbb{N}$ snapshots $\{y_{j_1}, \ldots, y_{j_{n_s}}\} \subset \{y_1, \ldots, y_K\}$ at the time steps $t_{j_1}, \ldots, t_{j_{n_s}}$ with indices $j_1, \ldots, j_{n_s} \in \{1, \ldots, K\}$. POD constructs orthonormal basis vectors $v_1, \ldots, v_n \in \mathbb{R}^n$ of the $n_r$-dimensional POD space that is the solution to the minimization problem

$$\min_{\v_1, \ldots, \v_n} \sum_{i=1}^{n_s} \left\| y_{j_i} - \sum_{l=1}^{n_r} (v_l^T y_{j_i}) v_l \right\|^2_2.$$  

The POD basis $V = (v_1, \ldots, v_{n_r})$ is formed of the left-singular vectors of the snapshot matrix $Y = (y_{j_1}, \ldots, y_{j_{n_s}}) \in \mathbb{R}^{n \times n_s}$ corresponding to the $n_r$ largest singular values.

The POD-Galerkin reduced model of (27) is

$$\hat{E}\hat{y}_i = \hat{A}\hat{y}_{i-1} + V^T f(V\hat{y}_{i-1}),$$  

where $\hat{y}_i \in \mathbb{R}^{n_r}$ is the reduced state vector at time step $t_i$ for $i = 1, \ldots, K$, and $\hat{E} = V^T EV, \hat{A} = V^T AV$ are the reduced operators.

Solving (28) requires evaluating the nonlinear function $f(V\hat{y}_{i-1})$ at the $n$-dimensional vector $V\hat{y}_{i-1} \in \mathbb{R}^n$, which can be computationally expensive. DEIM derives an approximation of $f(V\hat{y}_{i-1})$ to avoid evaluating $f$ at all $n$ components of $V\hat{y}_{i-1}$. To this end, DEIM constructs $p \in \mathbb{N}$ DEIM basis vectors $u^1, \ldots, u^p \in \mathbb{R}^n$ using POD on the nonlinear snapshots $f(y_{j_1}), \ldots, f(y_{j_{n_s}}) \in \mathbb{R}^n$. The DEIM basis vectors are the columns of the DEIM basis matrix $U = (u_1, \ldots, u_p) \in \mathbb{R}^{n \times p}$. Additionally, DEIM selects $p \in \mathbb{N}$ DEIM interpolation points $q_1, \ldots, q_p \in \{1, \ldots, n\}$ using a greedy strategy; see [22]. The DEIM mask matrix is $P = (e_{q_1}, \ldots, e_{q_p}) \in \{0, 1\}^{n \times p}$. The DEIM interpolant is the pair $(U, P)$. The DEIM approximation of the nonlinear function $f$ evaluated at the vector $V\hat{y}_i$ is given as

$$f(V\hat{y}_i) \approx U(P^T U)^{-1} P^T f(V\hat{y}_i).$$  

The POD-DEIM-Galerkin reduced model of (27) at a time step $t_i, i = 1, \ldots, K$, is

$$\hat{E}\hat{y}_i = \hat{A}\hat{y}_{i-1} + V^T U(P^T U)^{-1} P^T f(V\hat{y}_{i-1}).$$  

The reduced model (30) is often orders of magnitude faster to solve than the full model (27) and the reduced state vectors $\hat{y}_1, \ldots, \hat{y}_K \in \mathbb{R}^{n_r}$ lead to accurate approximations $V\hat{y}_1, \ldots, V\hat{y}_K \in \mathbb{R}^n$ of the full state vectors $y_1, \ldots, y_K \in \mathbb{R}^n$, respectively.

### 4.1.2. Online adaptive model reduction.

We adapt the DEIM interpolant of the nonlinear function $f$ in the online phase, i.e., we adapt the DEIM basis $U$ and the DEIM mask matrix $P$ during the time stepping. We proceed as follows. Let $U_0$ denote the DEIM basis matrix, which is derived using POD as discussed in subsection 4.1.1. Further, let $q_1^0, \ldots, q_p^0 \in \{1, \ldots, n\}$ be the DEIM interpolation points and $P_0 = (e_{q_1^0}, \ldots, e_{q_p^0})$ the mask matrix that are derived with the DEIM procedure in the offline phase; see subsection 4.1.1. Consider now the online phase at time step
to \( t_1 \). To compute the reduced state vector \( \tilde{y}_1 \), we first adapt the DEIM basis matrix \( U_0 \) and the mask matrix \( P_0 \) to \( U_1 \) and \( P_1 \), respectively, and then use the adapted DEIM interpolant \((U_1, P_1)\) in the reduced model (30) to compute the reduced state vector \( \tilde{y}_1 \). The DEIM basis matrix \( U_0 \) is adapted to \( U_1 \) using the GROUSE rank-one update, as we will discuss in detail in subsection 4.1.3. This process is continued iteratively, i.e., at time step \( t_i \), we adapt \( U_{i-1} \) and \( P_{i-1} \) to obtain \( U_i \) and \( P_i \), respectively, and then use the adapted interpolant \((U_i, P_i)\) for computing the reduced state vector \( \tilde{y}_i \) at time step \( t_i \). Note that the POD basis matrix \( V \) and the reduced linear operators \( \tilde{E} \) and \( \tilde{A} \) are kept unchanged online (although in principle they too could be adapted).

### 4.1.3. Subspace adaptation in online adaptive model reduction.

We use the GROUSE rank-one update with the residual-annihilating step size (22) to adapt the DEIM basis matrix. Consider time step \( t_i \) for \( i = 1, \ldots, K \). To adapt the DEIM basis matrix \( U_{i-1} \) to \( U_i \) at time step \( t_i \), we follow [43] and oversample the DEIM approximation. Let \( \{q_{i+1}, \ldots, q_{i+s}\} \subset \{1, \ldots, n\} \setminus \{q_1^{-1}, \ldots, q_{i-1}^{-1}\} \) be a set of \( s \in \mathbb{N} \) additional indices that are drawn uniformly from the set \( \{1, \ldots, n\} \setminus \{q_1^{-1}, \ldots, q_{i-1}^{-1}\} \), where \( q_1^{-1}, \ldots, q_{i-1}^{-1} \) are the DEIM interpolation points of the previous time step \( t_{i-1} \). The extended mask matrix \( S_i \in \{0, 1\}^{n \times m} \), \( m = p+s \), is assembled from the points in the set \( \{q_1^{-1}, \ldots, q_{i}^{-1}, q_{i+1}, \ldots, q_{i+p}\} \) as \( S_i = (e_{q_1^{-1}}, \ldots, e_{q_{i}^{-1}}, e_{q_{i+1}}, \ldots, e_{q_{i+p}}) \). The matrix \( S_i \) corresponds to \( m = p+s > p \) point indices, and therefore, the interpolation problem (29) of the classical DEIM approximation with the interpolant \((U_{i-1}, P_{i-1})\) becomes an overdetermined least-squares problem using the extended mask matrix \( S_i \)

\[
\alpha = \arg \min_{\tilde{\alpha} \in \mathbb{R}^p} \|S_i^T U_{i-1} \tilde{\alpha} - S_i^T f(V \tilde{y}_{i-1})\|_2^2
\]

with

\[
f(V \tilde{y}_{i-1}) \approx U_{i-1} \alpha.
\]

The solution \( \alpha \) of (31) is

\[
\alpha = (U_{i-1}^T S_i S_i^T U_{i-1})^{-1} U_{i-1}^T S_i S_i^T f(V \tilde{y}_{i-1}).
\]

The regression problem (31) fits into the framework of the GROUSE subspace adaptation approach of subsection 2.2, so that we can find the adapted DEIM basis matrix \( U_i \) with the low-rank update derived in Corollary 4. In addition to updating the DEIM basis matrix, the DEIM interpolation points \( q_1^{-1}, \ldots, q_{i-1}^{-1} \) are updated to \( q_1^i, \ldots, q_{i}^i \). For this task we use the algorithm introduced in [43, section 4]. The entire DEIM online adaptivity procedure is summarized in Algorithm 1.

### 4.1.4. Example of DEIM subspace adaptation.

We apply the online subspace adaptation to the DEIM interpolant of a reduced model of the FitzHugh–Nagumo system. The FitzHugh–Nagumo system is used in the original DEIM paper [22] as a benchmark example. The number of time steps is \( K = 10^6 \) and the dimension of the discretized state space is \( n = 2048 \). The state vectors \( y_0, y_{1000}, y_{2000}, \ldots, y_K \in \mathbb{R}^n \) at every \( 1000 \)th time step are used as snapshots to construct \( n_r = 10 \) POD basis vectors and the corresponding POD basis matrix \( V \in \mathbb{R}^{n \times n_r} \). The nonlinear function is evaluated at the snapshot time instances to obtain the nonlinear snapshots \( f(y(t_0)), f(y(t_{1000})), f(y(t_{2000})), \ldots, f(y(t_K)) \).

We compare the error of a static reduced model without online subspace adaptation to the error of an adaptive reduced model as in Algorithm 1. We report the average of the relative \( L_2 \) error of the approximation \( V \tilde{y}_i \) to the reference \( y_i \) at the
time steps \( t_{500}, t_{1500}, \ldots, t_{K-500} \). Thus, the error is measured at time steps other than where the snapshots were taken.

Figure 2(a) compares the \( L_2 \) error of the states of the reduced model (30) with a static DEIM interpolant to the error of the reduced model with an adaptive DEIM interpolation. The dimension of the DEIM subspace is varied over the range \( p \in \{2, 4, 6, 8, 10\} \). The DEIM subspace and the DEIM interpolation points are adapted every 50th time step, which means that we set \( l = 50 \) in Algorithm 1. At each adaptation step, the geometric rank-one update of Corollary 4 is performed to adapt the DEIM basis matrix based on \( s \), adaptation interval \( l \)

\[ \text{Algorithm 1. Time stepping a reduced model with online adaptive DEIM.} \]

\textbf{Input:} System matrices \( E, A \), nonlinear function \( f \), initial condition \( y_0 \), POD basis matrix \( V \), DEIM basis matrix \( U_0 \), DEIM interpolation points matrix \( P_0 \), number of sampling points \( s \), adaptation interval \( l \)

1. Set \( \hat{y}_0 = V^T y_0 \)
2. for \( i = 1, \ldots, K \) do
3. if \( \text{mod}(i, l) == 0 \) then
4. \{Adapt DEIM interpolant every \( l \)-th time step\}
5. Set \( q_{i-1}^{1}, \ldots, q_{p}^{1} \) to the interpolation points of \( P_{i-1} \)
6. Draw \( q_{p+1}^{i}, \ldots, q_{p+s}^{i} \) uniformly from \( \{1, \ldots, n\} \setminus \{q_{1}^{i-1}, \ldots, q_{p}^{i-1}\} \)
7. Construct mask matrix \( S_i \) from points \( q_{1}^{i-1}, \ldots, q_{p}^{i-1}, q_{p+1}^{i}, \ldots, q_{p+s}^{i} \)
8. Evaluate nonlinear function at sampling points \( b = S_i^T f(V \hat{y}_{i-1}) \)
9. \{Employ Corollary 4 to adapt \( U_{i-1} \}\}
10. Set \( \alpha = (U_{i-1}^TS_iS_i^TU_{i-1})^{-1}U_{i-1}^TS_iU_{i-1} \alpha \) and \( r = b - S_i^T U_{i-1} \alpha \)
11. Set \( v = \alpha/\|\alpha\|_2, s_1 = 2\|r\|_2/\|\alpha\|_2 \), and \( t^* = s_1^{-1} \arctan(\|r\|_2/\|\alpha\|_2) \)
12. Adapt basis matrix \( U_i = U_{i-1} + ((\cos(t^* s_1) - 1) U_{i-1} v + \sin(t^* s_1) (S_i r)/\|r\|_2) v^T \)
13. Adapt interpolation points matrix \( P_{i-1} \) to \( P_i \) with [43, Algorithm 2]
14. else
15. Set \( U_i = U_{i-1} \) and \( P_i = P_{i-1} \) \{No adaptation\}
16. end if
17. \( \tilde{f}_i = V^T U_i (P_i^TU_i)^{-1} P_i^T f(V \hat{y}_{i-1}) \) \{Approximate nonlinear function\}
18. Solve reduced model \( \tilde{E} \hat{y}_i = \tilde{A} \hat{y}_{i-1} + \tilde{f}_i \) for \( \hat{y}_i \)
19. end for

\textbf{Output:} Reduced states \( \hat{y}_0, \ldots, \hat{y}_K \).
derived in Corollary 4 and implemented in Algorithm 1. The curve with label “adapt, asym. optimal” corresponds to the step size \( \hat{t} = \frac{1}{s_1} \arcsin \left( \frac{\|r\|_2}{\|\alpha\|_2} \right) \) that is discussed in the GROUSE convergence analysis of [10]; see also Remark 5. We additionally compare to the constant step size 0.05 in “adapt, constant” and a decaying step size 0.05/i in “adapt, decaying step size,” as in, e.g., the GROUSE numerical experiments in [8], where i is the counter variable in the for-loop in Algorithm 1. The number of samples is set to \( s = 400 \) and the DEIM subspace and the DEIM interpolation points are adapted every 50th time step. The optimal and the asymptotically optimal step size lead to similar results (the curves are on top of each other), which was to be expected, since the functions \( \arctan \) and \( \arcsin \) match up to terms of third order. The less sophisticated choices “adapt, constant” and “adapt, decaying step size” lead to poor results which are even worse than those produced by the static subspace for DEIM basis dimensions of 8 and 10. This shows that for the application at hand, it is crucial to select a residual-related step size based on the ratio \( \frac{\|r\|_2}{\|\alpha\|_2} \), e.g., the minimizer \( \hat{t}^* \) from Corollary 4.

4.2. Subspace adaptation for gappy POD image reconstruction. In this section, the geometric subspace update is applied in combination with the method of gappy POD [27, 18] on an image processing problem, where we use the method to implant a new feature into a given subspace.

We briefly summarize gappy POD. Given a set of snapshots \( \{y_k\}_{k=1}^{n_s} \subset \mathbb{R}^n \), let \( \mathcal{U} = \text{colspan}(U) \) be the associated POD subspace represented by \( U \in \text{Sl}(n, p) \) with \( p \leq n_s \). Further, let \( y^g \in \mathbb{R}^n \) be an incomplete snapshot associated with an index set \( J = \{j_1, \ldots, j_m\} \subset \{1, \ldots, n\} \) of cardinality \( m \in \mathbb{N} \); \( y^g \) is incomplete in the sense that only components with indices in \( J \) are considered as accurate information. Gappy POD computes a vector contained in \( \mathcal{U} \) that best fits the incomplete snapshot \( y^g \) in a least-squares sense. Employing the mask matrix \( P = (e_{j_1}, \ldots, e_{j_m}) \in \{0, 1\}^{n \times m} \),

---

**Fig. 2.** The average relative \( L_2 \) error of a static reduced model is compared to the error of a reduced model with an online adaptive DEIM interpolant. The online adaptation based on the low-rank updates achieves up to an order of magnitude improvement in the \( L_2 \) error compared to the static DEIM interpolant.
the gappy POD approximation $y_{gpod} \in \mathbb{R}^n$ is determined by the masked least-squares minimization problem

$y_{gpod} = U\alpha_{gpod}, \quad \alpha_{gpod} = \arg \min_{\alpha \in \mathbb{R}^p} \| P^T U \alpha - P^T y^g \|^2.$

(Notice the similarities to the DEIM approach from subsection 4.1.4. Reference [28] exposes further details on the relation between gappy POD and the Empirical Interpolation Method (EIM, [11]), which predates DEIM.) In our concrete example of image processing, the snapshot set is taken from the so-called Yale Database [13]; see also [20, section 5.2].\footnote{More precisely, we have used row 11 of the set of 165 Yale images in (64 × 64)-MATLAB format provided by Deng Cai at http://www.cad.zju.edu.cn/home/dengcai/Data/FaceData.html.} Representing each image as a snapshot vector $y_k \in \mathbb{R}^n$, $n = 4096$, yields a snapshot matrix of dimension $Y \in \mathbb{R}^{4096 \times 10}$. The snapshots are displayed in Figure 4. The single image with glasses has been deliberately omitted from the snapshot set, so that no picture in the snapshot ensemble features the property “glasses-on.” The “glasses”-detail from this picture, displayed in the lower left corner of Figure 6, acts as a vector of gappy data $y^g \in \mathbb{R}^{4096}$ with $m = 1336$ nonzero entries and corresponding mask matrix $P$. The gappy POD objective is to find the linear combination of snapshots that comes closest to representing the “glasses”-feature in a least-squares sense. The resulting image is displayed in the second column of Figure 6 with the top picture showing the gappy POD solution and the bottom picture showing the reference image projected onto the subspace spanned by the POD modes. The
gappy POD reconstruction is a poor approximation of the reference picture because the POD space does not contain any information required to represent glasses.

\[ \| b - \Pi_P U^T b \| \]

Fig. 5. Plot of two periods of residual norm function (21) corresponding with the gappy POD subspace adaptation. The circle locates the first root and the star indicates the global maximum.

Fig. 6. Gappy POD approximation of a picture excerpt. To be read columnwise: Reference picture and training excerpt. Gappy POD reconstruction based on the excerpt and projection of complete reference onto the POD subspace. Gappy POD reconstruction using an adapted POD subspace and projection of complete reference thereon. Gappy POD reconstruction after adapting only the last column of POD subspace and projection of complete reference thereon.

Now, we use the GROUSE rank-one update combined with Corollary 4 to annihilate the gappy POD residual, which corresponds to solving the nonlinear equation (11) on \( Gr(n, p) = Gr(4096, 10) \). The input data are the mask matrix \( P \in \mathbb{R}^{n \times m} \) associated with the picture excerpt, the corresponding right-hand side \( b = P^T g \in \mathbb{R}^m \), and the subspace representative \( U_0 \in St(n, p) \) stemming from a POD of the input snapshots. A plot of the residual norm function along the rank-one update is displayed in Figure 5.

The update leads to a subspace representative \( U^* \) that allows for a perfect reproduction of the picture excerpt but also makes use of the information that was
previously sampled. We repeat the exercise while modifying only the last column of the initial POD subspace representative $U_0$ according to Corollary 10.

The gappy POD approximations using the adapted subspaces are shown in the last two columns of Figure 6, again in comparison with the projection of the reference image onto the respective subspace, as is clear from Corollary 4.\textsuperscript{6}

The important thing is how the adapted subspaces have changed. This can be visualized by projecting the initial snapshot ensemble onto the adapted subspaces; see Figure 7. Apart from the fact that the bright white spots in the original data set are reproduced in a graying way when projected onto the last-column adapted subspace, these two data sets look almost the same (Figure 7, bottom rows). In contrast, the original data set projected onto the fully adapted subspace features the property “glasses-on” throughout (Figure 7, top row). Nevertheless, the subspace distance between $[U_0]$ and the fully adapted $[U^*]$ is 0.1273, while the distance between $[U_0]$ and the subspace $[U^*]$ with only the last column adjusted is 1.2828, more than ten times as large. Recall from Remark 7 that the latter $[U^*]$ corresponds to an SVD update with respect to a column-replacement in the original subspace representative $U_0$.

Additional experiments are featured in section SM3 from the supplement. The supplement also includes MATLAB code for the adapted gappy POD examples discussed here.

5. Summary and conclusion. Subspace update problems arise in model reduction, machine learning, pattern recognition, and computer vision. This paper focuses on the particular use case of subspace adaptation in combination with the model reduction methods of gappy POD and DEIM. These methods have in common that a mask matrix is utilized to extract the features deemed most important to the underlying problem. In both cases, the objective of the downstream subspace adaptation is to produce subspaces that contain elements that match the selected components. We have formalized this objective as a nonlinear equation on the Grass-

\textsuperscript{6}This transfers in an analogous form to the sub-subspace setting of subsection 3.4 in which both reconstructed images coincide since they both correspond to copying the training set to the respective entries of the unmodified gappy POD solution.
mann manifold and have provided a closed-form solution that builds on the GROUSE approach [8, 49].

In the DEIM test case, discussed in subsection 4.1.4, the mask matrix operates on vectors contained in the subspace that represents the nonlinear terms of the underlying discretized PDE. In the gappy POD test cases, discussed in subsection 4.2, the mask matrix selects the important components from vectors contained in the subspace of state vector solution candidates. In the test case of DEIM-based model reduction, the Grassmann subspace update is used as an online adaptation method to improve the fit of the components sampled from the nonlinear term. The reduced model with online subspace updating achieves an average error of about one order of magnitude lower than a classical reduced model without the adaptation. In the gappy POD image processing example, the Grassmann subspace update is applied to implement a new feature in the subspace of solution candidates that is not contained in the sample data set. We expect the method to show similar advantages when used in combination with the missing point estimation [7, 51], because of the similarities to DEIM and gappy POD.

Appendix A. A direct solution of the Grassmann residual equation (11). This appendix features a short solution of (11). Obviously, (11) is solved if we can find \( t^* \in \mathbb{R} \) and \( \alpha^* \in \mathbb{R}^p \) such that \( P^T U(t^*) \alpha^* = b \), where \( U(t^*) := U_0 + (\cos(t^* s_1) - 1) U_0 v + \sin(t^* s_1) \frac{v}{||v||_2} ) v^T \). (All occurring quantities to be understood were introduced in Theorem 3.) Mind that \( v = \alpha/||\alpha||_2 \). Using an additional real parameter \( \lambda \) and the ansatz \( \alpha^* = \lambda \alpha = \lambda v ||\alpha||_2 \) leads to the equation

\[
\lambda \cos(t^* s_1) \left( 1 - \tan(t^* s_1) \frac{||\alpha||_2}{||r||_2} \right) P^T U_0 \alpha + \tan(t^* s_1) \frac{||\alpha||_2}{||r||_2} b = b.
\]

By setting \( t^* = \frac{1}{s_1} \arctan \left( \frac{||r||_2}{||\alpha||_2} \right) \), the terms involving \( P^T U_0 \) cancel, which leaves an equation for \( \lambda \):

\[
\lambda \left( \arctan \left( \frac{||r||_2}{||\alpha||_2} \right) \right) b = b.
\]

The solution is \( \lambda = \frac{1}{\cos(\arctan \left( \frac{||r||_2}{||\alpha||_2} \right) \right)} = \sqrt{1 + \frac{||r||_2^2}{||\alpha||_2^2}} \).

In addition to its concision, this approach has the advantage that it simultaneously gives both \( t^* \) and the associated vector of coefficients \( \alpha^* = \sqrt{\left( \frac{||r||_2^2}{||\alpha||_2^2} + 1 \right)} \alpha \in \mathbb{R}^p \). On the other hand it does not allow one to keep track of the residual depending on \( t \), because for \( t \neq t^* \), a defining equation is missing and \( \alpha(t) \) and \( \alpha \) are not collinear.

Nevertheless, we remark that the above shortcut approach may be adapted to apply also in the setting of Corollary 10 from subsection 3.4. In this case, one can work from the ansatz \( \alpha^* = (\alpha_1, \ldots, \alpha_{p-I}, \lambda(\alpha_{p-I+1}, \ldots, \alpha_p))^T \).

One may also start by first applying the orthogonal coordinate transformation \( \Phi = (v, Z) \in O_p \) to the subspace representative \( U_0 \), where \( Z \in \mathbb{R}^{p \times (p-1)} \) contains an arbitrary orthonormal basis of \( v^* \), and then work with \( U_0 \Phi, U(t^*) \Phi \). This course of action essentially leads to (33) appearing in the first column of \( U(t^*) \Phi \) and the rest of the argument is analogous. See [49, App. C, Proof of Lemma 4] for related considerations.

Appendix B. Addendum to subsection 3.4. A simple example of a differentiable Grassmann objective function for which Proposition 9 does not hold is

\[
f : Gr(n, p) \to \mathbb{R}, \quad [U] \mapsto x^T U U^T y,
\]
where \( x, y \in \mathbb{R}^n \) are not orthogonal to \([U]\).

By using the basic fact that \( D_X(v^TXw) = (\frac{\partial}{\partial x_{ij}} v^T X w)_{ij} = vw^T \) and the product rule, we see that the Grassmann gradient is

\[
\nabla_{[U]} f = (I - UU^T)D_U f = (I - UU^T)(xy^T + yx^T)U,
\]

where \( D_U f = (\frac{\partial f}{\partial u_{ij}})_{ij} \in \mathbb{R}^{n \times p} \); see [25, eq. (2.70)]. (Note that \( \nabla_{[U]} f \) is of rank two in general, but of rank one if \( x = y \).) Introducing \( U = (U_1, U_2) \) with \( U_1 \in St(n, p - l) \), \( U_2 \in St(n, l) \), we may write \( UU^T = U_1U_1^T + U_2U_2^T \). By fixing \( U_1 \), \( f \) becomes a function \( f_2 : Gr(n, l) \rightarrow \mathbb{R}, [U_2] \rightarrow x^TU_1U_1^Ty + x^TU_2U_2^Ty \). The gradient is

\[
\nabla_{[U_2]} f_2 = (I - U_2U_2^T)(xy^T + yx^T)U_2 \in \mathbb{R}^{n \times l}.
\]

Likewise, for \( f_1 : Gr(n, p - l) \rightarrow \mathbb{R}, [U_1] \rightarrow x^TU_1U_1^Ty + x^TU_2U_2^Ty \), we obtain

\[
\nabla_{[U_1]} f_1 = (I - U_1U_1^T)(xy^T + yx^T)U_1 \in \mathbb{R}^{n \times l}.
\]

Splitting up the original gradient into an \((n \times (p - l))\) and an \((n \times l)\) matrix gives

\[
\nabla_{[U]} f = ((I - UU^T)(xy^T + yx^T)U_1, (I - UU^T)(xy^T + yx^T)U_2) \\
\neq ((I - U_1U_1^T)(xy^T + yx^T)U_1, (I - U_2U_2^T)(xy^T + yx^T)U_2) \\
= (\nabla_{[U_1]} f_1, \nabla_{[U_2]} f_2).
\]

In particular, \( U_1^T \nabla_{[U_2]} f_2 = U_1^T xy^TU_1 + U_1^T yx^TU_2 \neq 0 \) and the geodesic \( U_2(t) \) in \( Gr(n, l) \) along the gradient direction \( \nabla_{[U_2]} f_2 \) is not orthogonal to \( U_1 \),

\[
U_1^T U_2(t) \neq 0.
\]

A sufficient condition for (25) and Proposition 9 to hold is \( (I - UU^T)D_U f = D_U f \) or, in short, \( U^TD_U f = 0 \).

Acknowledgments. The authors would like to thank Laura Balzano (University of Michigan) for helpful conversations concerning GROUSE. Moreover, the authors thank the reviewers and Associate Editor for inspiring comments that improved this manuscript.

REFERENCES


SUBSPACE UPDATES FOR ADAPTIVE MOR


