Communications in Information and Systems
Vol. 8, No. 2, pp. 71-84, 2008

001

NUMERICAL REPRESENTATIONS OF A UNIVERSAL SUBSPACE FLOW FOR LINEAR PROGRAMS*

P.-A. ABSIL†

Abstract. In 1991, Sonnevend, Stoer, and Zhao [Math. Programming 52 (1991) 527–553] have shown that the central paths of strictly feasible instances of linear programs generate curves on the Grassmannian that satisfy a universal ordinary differential equation. Instead of viewing the Grassmannian $\mathrm{Gr}(m,n)$ as the set of all $n\times n$ projection matrices of rank m, we view it as the set $\mathbb{R}_*^{n\times m}$ of all full column rank $n\times m$ matrices, quotiented by the right action of the general linear group $\mathrm{GL}(m)$. We propose a class of flows in $\mathbb{R}_*^{n\times m}$ that project to the flow on the Grassmannian. This approach requires much less storage space when $n \gg m$ (i.e., there are many more constraints than variables in the dual formulation). One of the flows in $\mathbb{R}_*^{n\times m}$, that leaves invariant the set of orthonormal matrices, turns out to be a particular version of a matrix differential equation known as Oja's flow. We also point out that the flow in the set of projection matrices admits a double bracket expression.

Keywords: Linear programming, Grassmannian, Grassmann manifold, Stiefel manifold, ordinary differential equation, Oja's flow, double bracket flow.

Mathematics Subject Classification: 37N40 Dynamical systems in optimization and economics; 90C05 Linear programming

1. Introduction. One of Roger Brockett's major contributions to date has been to propose and analyze the matrix differential equation [Bro91]

$$H'(t) = [H(t),[H(t),N]], \quad H(0) = H_0, \tag{1}$$

where N and H_0 are $n \times n$ real symmetric matrices and $[A,B] := AB - BA$ denotes the matrix commutator. This matrix flow belongs to a class of flows on manifolds that realize computational algorithms. It is able to solve the eigenvalue problem of a symmetric matrix A (see [Bro91], or [HM94, §2.1], [Deh95, §4.2]): to this end, choose $H_0 = A$ and $N = \mathrm{diag}(\mu_1,\ldots,\mu_n)$ with $\mu_1 > \cdots > \mu_n$; then $\lim_{t\to\infty H(t)} H(t)$ exists and is a diagonal matrix, whose diagonal elements are the eigenvalues of A since the flow is isospectral (i.e., the spectrum of H does not vary along the trajectory). The differential equation (1) is also capable of sorting lists: if N is chosen, for example, as $\mathrm{diag}(1,2,\ldots,n)$, then for almost all orthogonal $n\times n$ matrices Θ and for $H(0) = \Theta^T[\mathrm{diag}(\lambda_1,\lambda_2,\ldots,\lambda_n)]\Theta$, the solution of (1) will approach $H(\infty) = \mathrm{diag}(\lambda_{\pi(1)},\lambda_{\pi(2)},\ldots,\lambda_{\pi(n)})$ with the final list sorted by size. Brockett also

*Dedicated to Roger Brockett on the occasion of his 70th birthday. This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. The scientific responsibility rests with its authors.

†Département d'ingénierie mathématique, Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium (www.inma.ucl.ac.be/∼absil).

observed that this sorting property can be exploited to solve linear programming problems when the feasible set is bounded and the vertices are known. In view of its many remarkable properties, the flow (1), known as the *double bracket* flow (sometimes affectionately dubbed the "double Brockett" flow) has become a center of interest in the numerical integration, automatic control, and linear algebra communities; see, e.g., [CD91, BBR92, HM94, Deh95, Chu95, CIZ97, FO01, Ise02, Tam04, BI04, Cas04].

Also in 1991, Sonnevend, Stoer, and Zhao [SSZ91, §4] have proposed a universal flow related to the central path of linear programs. They consider linear programs in standard primal-dual form, with primal

$$\begin{aligned} \min \quad & c^T x \\ s.t. \quad & Ax = b \\ & x \geq 0 \end{aligned} \tag{2}$$

and dual

$$\begin{aligned} \max \quad & b^T y \\ s.t. \quad & A^T y + s = c \\ & s \geq 0 \end{aligned} \tag{3}$$

where $A \in \mathbb{R}^{m \times n}$ is of full row rank, and b, c, x, and s are vectors of appropriate dimensions. The universal flow takes the form (see Section 6 for details)

$$M' = M\lfloor M\mathbf{1}\rceil + \lfloor M\mathbf{1}\rceil M - 2M\lfloor M\mathbf{1}\rceil M, \tag{4}$$

where $M(t) \in \mathbb{R}^{n \times n}$ is an orthogonal projector with rank m for all t (i.e., $M(t)$ is symmetric, idempotent, of rank m for all t), $\mathbf{1}$ is the vector of all ones, and $\lfloor v \rceil$ denotes the diagonal matrix of vector v. An analysis of the flow (4) on the set of projectors, comprising a characterization of the eigenvectors and eigenvalues of the linearization at the equilibria, was carried out by Zhao [Zha08].

In [SSZ91, §4], it is also pointed out that the differential equation (4) admits the more symmetric form

$$M' = MD - DM, \quad \text{where } D = M\lfloor M\mathbf{1}\rceil - \lfloor M\mathbf{1}\rceil M. \tag{5}$$

In other words, (5) reads

$$M' = [M, [M, \lfloor M\mathbf{1}\rceil]]. \tag{6}$$

This is the double bracket flow (1) where N, instead of being constant, depends on the state of the dynamical system.

It was the simplest but perhaps most noticeable goal of this paper to draw the reader's attention to this interesting connection between the double bracket flow (1) and the universal flow for linear programs (4), a fact that seemingly has been overlooked ever since the two flows were proposed in 1991. Another, less straightforward contribution of this paper is to provide a self-contained yet rather concise derivation of various representations of the universal flow (4). Instead of going first for the flow in $\mathbb{R}^{n\times n}$, we start by deriving a flow in $\mathbb{R}^{n\times m}$ of the form

$$Z' = F(Z), \tag{7}$$

that projects to (4) through the operation $M = Z(Z^TZ)^{-1}Z^T$. The expression (7) evolves in a space of dimension nm, to be compared with (4) that involves n^2 variables. Arguably, (7) is more tractable than (4) for numerical computation when $n \gg m$, i.e., there are many more constraints than variables in the dual (3). We also derive a flow in $\mathbb{R}^{n\times m}$ that leaves the set of orthonormal matrices invariant and projects to (4). This flow can be viewed as a special instance of Oja's flow [Oja82, Oja89], and its dynamics does not involve matrix inversion.

This paper is organized as follows. Section 2 gives the necessary background in linear programming. In Section 3, we derive a universal flow related to the central path of linear programs. This flow is shown to induce a subspace flow in Section 4. The orthonormal version is derived in Section 5. The link with (4) is made in Section 6. Final remarks are made in Section 7.

2. Preliminaries on linear programming and the central path. In this section, we briefly review the basics of linear programming and the central path. We refer the reader, e.g., to Wright [Wri97] or Ye [Ye97] for more details. We also introduce some assumptions on the linear programs.

For any vectors $x, s \in \mathbb{R}^n$ and scalar $\alpha \in \mathbb{R}$, we denote $x \circ s = (x_1s_1, \ldots, x_ns_n)^T$ and $x^\alpha = (x_1^\alpha, \ldots, x_n^\alpha)$. We let $\lfloor x \rceil$ denote the diagonal matrix of vector x and $\mathbf{1}$ denote the vector of all ones in $\mathbb{R}^n$. For a matrix $Q \in \mathbb{R}^{j\times k}$, Q_{JK} stands for the submatrix of Q consisting of all entries of Q with row indices in the index set J and all column indices in the index set K. For ease of notation, we write A_K for the submatrix of the coefficient matrix A consisting of all entries with column indices in K.

Consider the linear program in standard primal-dual form (2)–(3), with $A \in \mathbb{R}^{m\times n}$ of full row rank, $b \in \mathbb{R}^m$, and $c \in \mathbb{R}^n$. The vector (x^*, y^*, s^*) is a primal-dual solution for (2)–(3) if and only if it satisfies the Karush-Kuhn-Tucker (KKT) conditions

$$x \circ s = 0 \tag{8a}$$

$$Ax = b \tag{8b}$$

$$A^Ty + s = c \tag{8c}$$

$$x \geq 0, \quad s \geq 0 \tag{8d}$$

We define the primal-dual feasible set $\mathcal{F}$ and strictly feasible set $\mathcal{F}^o$ by

$$\mathcal{F} = \{(x, y, s) : Ax = b, A^T y + s = c, (x, s) \geq 0\}, \tag{9a}$$

$$\mathcal{F}^o = \{(x, y, s) : Ax = b, A^T y + s = c, (x, s) > 0\}, \tag{9b}$$

and we use Ω_P and Ω_D to denote the primal and dual solution sets, i.e.,

$$\Omega_P = \{x^* : x^* \text{ solves } (2)\}, \qquad \Omega_D = \{(y^*, s^*) : (y^*, s^*) \text{ solves } (3)\}.$$

We assume that the primal and dual are strictly feasible (i.e., $\mathcal{F}^o$ is nonempty), which ensures that the primal solution set Ω_P is nonempty and bounded and that the set

$$\{s^* : (y^*, s^*) \in \Omega_D \text{ for some } y^* \in \mathbb{R}^m\}$$

is nonempty and bounded.

We can define two index sets $\mathcal{B}$ and $\mathcal{N}$ as follows:

$$\mathcal{B} = \{j \in \{1, 2, \ldots, n\} : x_j^* \neq 0 \text{ for some } x^* \in \Omega_P\}, \tag{10a}$$

$$\mathcal{N} = \{j \in \{1, 2, \ldots, n\} : s_j^* \neq 0 \text{ for some } (y^*, s^*) \in \Omega_D\}. \tag{10b}$$

The Goldman-Tucker theorem guarantees that every $j \in \{1, 2, \ldots, n\}$ belongs to either $\mathcal{B}$ or $\mathcal{N}$ but not both. The index set $\mathcal{B}$ is termed the *optimal basis* of (A, b, c). Observe that $\mathcal{B} = \{j \in \{1, 2, \ldots, n\} : s_j^* = 0 \text{ for all } (y^*, s^*) \in \Omega_D\}$. Hence $\mathcal{B}$ specifies the constraints in the dual that are active at all the solutions. From the knowledge of $\mathcal{B}$, it is straightforward to deduce the solutions of the linear program: $\Omega_P = \{x : A_\mathcal{B} x_\mathcal{B} = b,\ x_\mathcal{B} \geq 0,\ x_\mathcal{N} = 0\}$, $\Omega_D = \{(y, s) : c - A^T y = s,\ s_\mathcal{B} = 0,\ s_\mathcal{N} \geq 0\}$. In particular, in the nondegenerate case where there is only one primal and one dual solution, the cardinality of $\mathcal{B}$ is m and y^* is the solution of $(A_\mathcal{B})^T y = c$.

The universal continuous-time flows that we consider in this paper are able to produce $\mathcal{B}$ when t goes to infinity, for initial conditions adequately chosen as a function of (A, b, c). Hence they are potentially useful for solving linear programming problems.

The *central path* is the curve $t \mapsto (x(t), s(t))$ defined by

$$x \circ s = e^{-t}\mathbf{1} \tag{11a}$$

$$Ax = b \tag{11b}$$

$$A^T y + s = c \tag{11c}$$

$$x > 0, \quad s > 0. \tag{11d}$$

Under our strict feasibility assumption, (11) has one and only one solution $(x(t), s(t))$ for all t. Moreover, $x^* := \lim_{t\to\infty} x(t)$ and $s^* := \lim_{t\to\infty} s(t)$ exist and satisfy

$$x_\mathcal{B}^* > 0,\ x_\mathcal{N}^* = 0,\ s_\mathcal{B}^* = 0,\ s_\mathcal{N}^* > 0,$$

where $(\mathcal{B}, \mathcal{N})$ is the optimal basis of (A, b, c). Hence the optimal basis can be deduced from the limit of the central path.

From now on, we assume that A has full row rank. (Any problem that does not satisfy this assumption can be transformed to an equivalent problem that does.)

3. A universal flow derived from the central path. We are interested in finding matrix-valued expressions Γ and F such that, for all strictly feasible linear programs (2)–(3) with coefficients (A, b, c), it holds that

$$\frac{\mathrm{d}}{\mathrm{d}t}\Gamma(A, x(t), s(t)) = F(\Gamma(A, x(t), s(t))), \tag{12}$$

where $t \mapsto (x(t), s(t))$ is the central path of (A, b, c). The operations allowed in the expressions Γ and F are matrix addition, multiplication, inversion, square root, and the diagonal operation $u \mapsto \lfloor u \rceil$, involving the arguments passed to the expression as well as the vector of all ones $\mathbf{1}$ and constant scalars. (Observe that submatrix extraction is not an allowed operation and that the matrix $\Gamma(A, x(t), s(t))$, but not A, $b(t)$, $c(t)$, is passed to the expression F in (12).) Moreover, these expressions must be such that the optimal basis $(\mathcal{B}, \mathcal{N})$ of the linear program can be easily deduced from $\lim_{t\to\infty} \Gamma(A, x(t), s(t))$.

Expressions Γ and F satisfying these properties can be exploited, at least theoretically, to solve linear programming problems, according to the following procedure: (i) Compute a point $(x(\tau), s(\tau))$ on the central path for some τ. (ii) Evaluate

$$Z_0 := \Gamma(A, x(\tau), s(\tau)).$$

(iii) Integrate the universal flow

$$\frac{\mathrm{d}}{\mathrm{d}t}Z = F(Z), \quad Z(0) = Z_0, \tag{13}$$

assumed to have exactly one solution trajectory $t \mapsto Z(t)$. (iv) Evaluate $\lim_{t\to\infty} Z(t)$ and deduce the optimal basis $(\mathcal{B}, \mathcal{N})$. Even if this procedure is not competitive with state-of-the-art linear programming solvers, its analysis can shed light on the partition of the set of all linear programs into subsets that share the same optimal basis.

The flow (4) studied by Zhao [Zha08] fits in this framework; see Section 6 for details. The version of (13) that we derive in the present section is closely related to (4), as we will show in Section 6.

In order to achieve the goal presented above, let us take the derivative of the perturbed KKT conditions (11) with respect to time. This yields

$$x \circ s' + s \circ x' = -x \circ s \qquad (= -e^{-t}\mathbf{1}) \tag{14a}$$

$$Ax' = 0 \tag{14b}$$

$$A^T y' + s' = 0. \tag{14c}$$

The following identities will be useful. If h is a vector and $\Lambda, \tilde{\Lambda}$ are diagonal matrices of compatible dimensions, then

$$\lfloor h \rceil \mathbf{1} = h$$
$$\lfloor \Lambda \mathbf{1} \rceil = \Lambda$$
$$\lfloor \Lambda h \rceil = \Lambda \lfloor h \rceil$$
$$\Lambda \tilde{\Lambda} = \tilde{\Lambda} \Lambda.$$

We define the diagonal matrix

$$D(t) = \lfloor x^{1/2}(t) \circ s^{-1/2}(t) \rceil \tag{15}$$

and the projection matrix

$$M(t) = D(t) A^T (A D^2(t) A^T)^{-1} A D(t). \tag{16}$$

The projection matrix $M(t)$ is well defined for all $t \geq 0$. Indeed, for all $t \geq 0$, $x(t)$ and $s(t)$ are positive, hence $AD(t)$ is of full row rank, and thus $(AD^2(t)A^T)$ is invertible.

From the perturbed complementarity condition (11a), we have

$$D(t) = e^{t/2} \lfloor x(t) \rceil = e^{-t/2} \lfloor s^{-1}(t) \rceil \tag{17}$$

and thus

$$D(t)\mathbf{1} = e^{t/2} x(t) = e^{-t/2} s^{-1}(t).$$

This also yields

$$x(t) = e^{-t/2} D(t) \mathbf{1} \tag{18}$$

$$s(t) = e^{-t/2} D^{-1}(t) \mathbf{1} \tag{19}$$

and

$$x'(t) = -e^{-t/2} \left(\frac{1}{2} D(t) - D'(t) \right) \mathbf{1} \tag{20}$$

$$s'(t) = -e^{-t/2} D^{-2}(t) \left(\frac{1}{2} D(t) + D'(t) \right) \mathbf{1}. \tag{21}$$

We now take into account the derivatives of the primal and dual equality constraints, i.e., (14b) and (14c). Replacing (20) in (14b) yields

$$A(\frac{1}{2} D - D') \mathbf{1} = 0. \tag{22}$$

Note that there are more variables in D (n) than equations in (22) (m). Replacing (21) in (14c) and taking (22) into account yields

$$A^T y' = e^{-t/2} D^{-2} (\frac{1}{2} D + D') \mathbf{1}$$
$$A D^2 A^T y' = e^{-t/2} A (\frac{1}{2} D + D') \mathbf{1}$$
$$A D^2 A^T y' = e^{-t/2} A D \mathbf{1}$$

and finally

$$y' = e^{-t/2}(AD^2A^T)^{-1}AD\mathbf{1}.$$

Replacing this last result in (14c) yields

$$s' = -e^{-t/2}A^T(AD^2A^T)^{-1}AD\mathbf{1}.$$

From this equation and from (21), one has

$$D'\mathbf{1} = DM\mathbf{1} - \frac{1}{2}D\mathbf{1}$$

or equivalently

$$D' = D\lfloor M\mathbf{1}\rceil - \frac{1}{2}D = \lfloor M\mathbf{1}\rceil D - \frac{1}{2}D.$$

It remains to multiply this equation on the right by A^T to obtain

$$D'A^T = \lfloor DA^T(ADDA^T)^{-1}AD\mathbf{1}\rceil DA^T - \frac{1}{2}DA^T.$$

In view of the expression (15) of D as a function of x and y, the latter equation takes the universal ODE form (12) with

$$\Gamma(A, x, s) := \lfloor x^{1/2}\rceil\lfloor s^{-1/2}\rceil A^T \quad (= DA^T), \tag{23}$$

and

$$F(Z) := \lfloor Z(Z^TZ)^{-1}Z^T\mathbf{1}\rceil Z - \frac{1}{2}Z. \tag{24}$$

We record this result in the following proposition.

Proposition 1. *Let* (A, b, c) *define a strictly feasible linear program* (2)–(3) *with* A *of full row rank, and let* $t \mapsto (x(t), s(t))$ *be the central path of the linear program as defined by* (11)*. Then the universal ODE* (12)*, i.e.,*

$$\frac{\mathrm{d}}{\mathrm{d}t}\Gamma(A, x(t), s(t)) = F(\Gamma(A, x(t), s(t))),$$

holds with Γ *and* F *defined as in* (23) *and* (24)*.*

Note that the goal set in the beginning of this section is not yet achieved: we still need to show that the optimal basis $(\mathcal{B}, \mathcal{N})$ can be deduced from $\lim_{t\to\infty}\Gamma(A, x(t), y(t))$. This will follow in Section 7 from the relation between the flow of F (24) and (4) discussed in Section 6, and from the results of Zhao [Zha08] on the limit behavior of (4).

Even though they will not be used in the rest of the paper, we find it interesting to point out the following relations, which can also be found in [SSZ91, p. 534]:

$$x^{-1} \circ x' = -(I - M)\mathbf{1}$$
$$s^{-1} \circ s' = -M\mathbf{1}.$$

They can be derived from

$$x' = -x + x \circ s^{-1} \circ A^T y'$$
$$y' = (AD^2A^T)^{-1}Ax$$
$$s' = -A^T(AD^2A^T)^{-1}Ax$$

which follow from (14).

4. Grassmannian dynamics on the noncompact Stiefel manifold. We now study the flow of the vector field F (24), i.e.,

$$Z' = \lfloor Z(Z^TZ)^{-1}Z^T\mathbf{1}\rceil Z - \frac{1}{2}Z. \tag{25}$$

The right-hand side is well defined as long as Z has full column rank. Observe that (25) takes the form

$$Z' = \Lambda(Z)Z$$

where

$$\Lambda(Z) := \lfloor Z(Z^TZ)^{-1}Z^T\mathbf{1}\rceil - \frac{1}{2}I \tag{26}$$

is well-defined for all full-rank Z and diagonal. Therefore, we can write the solution of (25) as

$$Z(t) = \tilde{D}(t)Z_0,$$

where $Z_0 = Z(0)$ and $\tilde{D}(t)$ satisfies

$$\tilde{D}' = \left(\lfloor \tilde{D}Z_0(Z_0^T\tilde{D}^2Z_0)^{-1}Z_0^T\tilde{D}\mathbf{1}\rceil - \frac{1}{2}I\right)\tilde{D}, \quad \tilde{D}(0) = I.$$

Since $\tilde{D}Z_0(Z_0^T\tilde{D}^2Z_0)^{-1}Z_0^T\tilde{D}$ is an orthogonal projector, its elements remain bounded and thus the diagonal elements of

$$\tilde{D}(t) = \exp\left(\int_0^t \left(\lfloor \tilde{D}(\tau)Z_0(Z_0^T\tilde{D}^2(\tau)Z_0)^{-1}Z_0^T\tilde{D}(\tau)\mathbf{1}\rceil - \frac{1}{2}I\right)d\tau\right)$$

are nonzero for all t. This shows that, if $Z(0)$ is full rank, then (25) admits exactly one solution $Z(t)$ for all t, and $Z(t)$ is full rank for all t. In other words, (25) defines a flow on the *noncompact Stiefel manifold*

$$\mathbb{R}_*^{n\times m} = \{Z \in \mathbb{R}^{n\times m} : \text{rank}(Z) = m\}.$$

Note that the vector field F (24) satisfies the property

$$F(ZR) = F(Z)R$$

for all $Z \in \mathbb{R}_*^{n\times m}$ and all $m\times m$ invertible matrices R. Hence the vector field F defined in (24) induces a flow on the Grassmann manifold $\mathrm{Gr}(m,n)$ of m-planes in $\mathbb{R}^n$, as a direct consequence of [ASM08, Th. 3.4]. In other words, we have the property stated in Proposition 2 below. Let

$$\mathsf{span} : \mathbb{R}_*^{n\times m} \to \mathrm{Gr}(m,n) : Z \mapsto \{Zv : v \in \mathbb{R}^m\}$$

denote the column space mapping.

PROPOSITION 2. *Let $t \mapsto Z(t)$, resp. $t \mapsto \tilde{Z}(t)$, denote the solution trajectory of* (25) *with initial condition $Z(0) = Z_0 \in \mathbb{R}_*^{n\times m}$, resp. $\tilde{Z}(0) = \tilde{Z}_0 \in \mathbb{R}_*^{n\times m}$. If $\mathsf{span}(Z_0) = \mathsf{span}(\tilde{Z}_0)$, then $\mathsf{span}(Z(t)) = \mathsf{span}(\tilde{Z}(t))$ for all t.*

This property means that (25) induces a subspace flow. Note that two matrices Z and $\tilde{Z}$ in $\mathbb{R}_*^{n\times m}$ satisfy $\mathsf{span}(Z) = \mathsf{span}(\tilde{Z})$ if and only if there is an invertible $m\times m$ matrix R such that $\tilde{Z} = ZR$.

It also follows from [ASM08, Th. 3.4] that the matrix differential equation

$$Z' = \Lambda(Z)Z + ZU(Z), \tag{27}$$

where Λ is as in (26) and U is a continuously differentiable function on $\mathbb{R}_*^{n\times m}$ into $\mathbb{R}^{m\times m}$, induces the same subspace flow, regardless of U (under an existence and uniqueness condition). For $U(Z) = \frac{1}{2}I$, (27) becomes

$$Z' = C(Z)Z$$

with $C(Z) = \lfloor Z(Z^TZ)^{-1}Z^T\mathbf{1}\rceil$, which clearly reveals that the subspace flow induced by (27)—and by (25) in particular—is the *power flow* of the nonconstant matrix $C(Z)$; see, e.g., [ASM08, §4].

5. Dynamics on the orthogonal Stiefel manifold. It is possible to choose the function U in (27) such that the *orthogonal Stiefel manifold*

$$\mathrm{St}(m,n) = \{Z \in \mathbb{R}^{n\times m} : Z^TZ = I\}$$

is invariant for the flow of (27). To this end, observe that, from (27), and for $Z \in \mathrm{St}(m,n)$, we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(Z^TZ) &= Z^T\Lambda(Z)Z + U(Z)^TZ^TZ + Z^T\Lambda(Z)Z + Z^TZU(Z)\\ &= 2Z^T\Lambda(Z)Z + U(Z) + U(Z)^T.\end{aligned}$$

Hence, to enforce invariance of $\mathrm{St}(m,n)$, we set $U(Z) := -Z^T\Lambda(Z)Z$, and (27) becomes $Z' = (I - ZZ^T)\Lambda(Z)Z$. Observe that, for $Z \in \mathrm{St}(m,n)$, $\Lambda(Z) = \lfloor ZZ^T\mathbf{1}\rceil - \frac{1}{2}I$, which yields $Z' = (I - ZZ^T)(\lfloor ZZ^T\mathbf{1}\rceil - \frac{1}{2}I)Z$. Since $(I - ZZ^T)Z = 0$ for all $Z \in \mathrm{St}(m,n)$, (27) finally takes the form

$$Z' = (I - ZZ^T)\lfloor ZZ^T\mathbf{1}\rceil Z, \tag{28}$$

$Z \in \mathrm{St}(m,n)$. We summarize this result in the following proposition.

PROPOSITION 3. *Let $t \mapsto Z(t)$, resp. $t \mapsto \tilde{Z}(t)$, denote the solution trajectories of* (25)*, resp.* (28)*, such that $Z(0) = Z_0 \in \mathbb{R}_*^{n\times m}$, resp. $\tilde{Z}(0) = \tilde{Z}_0 \in \mathrm{St}(m,n)$. Then $\tilde{Z}(t) \in \mathrm{St}(m,n)$ for all t. Moreover, if $\mathsf{span}(Z_0) = \mathsf{span}(\tilde{Z}_0)$, then $\mathsf{span}(Z(t)) = \mathsf{span}(\tilde{Z}(t))$ for all t.*

The orthogonal Stiefel manifold $\mathrm{St}(m,n)$ is invariant by (28), but it is not asymptotically stable. To enforce its stability while preserving the induced subspace flow, it is possible to add a term (akin to the one in [MHM05]), as follows:

$$Z' = (I - ZZ^T)\lfloor ZZ^T\mathbf{1}\rceil Z + \mu Z(I - Z^T Z), \tag{29}$$

where $\mu > 0$ is chosen "sufficiently large" (we will be more specific in a moment). Using [Deh95, Lemma 3.9], we obtain that the singular values $\sigma_i(t)$, $i = 1,\dots,m$, of $Z(t)$ satisfy

$$\sigma_i'(t) = (e_i^T U^T C(Z(t)) U e_i + \mu)(1 - \sigma_i^2(t))\sigma_i(t),$$

where $Z = U\Sigma V^T$ is a singular value decomposition and $C(Z) = \lfloor ZZ^T\mathbf{1}\rceil$. Simple counterexamples show that there is some $Z \in \mathrm{St}(m,n)$ such that $e_i^T U^T C(Z(t)) U e_i < 0$; hence $\sigma_i = 1$ is not asymptotically stable when $\mu = 0$, i.e., the orthogonal Stiefel manifold $\mathrm{St}(m,n)$ is not asymptotically stable. However, if μ is chosen such that there is $\epsilon > 0$ for which $e_i^T U^T C(Z(t)) U e_i + \mu > \epsilon$ for all t, then all the singular values go monotonically to 1 as t goes to infinity, which means that $Z(t)$ converges to $\mathrm{St}(m,n)$. To choose μ, one can observe that $e_i^T U^T C(Z) U e_i \le \|U^T C(Z) U\|_2 = \|C(Z)\|_2 = \|\lfloor U\Sigma^2 U^T \mathbf{1}\rceil\|_2 \le \max_{\|u\|=1,\|v\|=\sqrt{n}} u^T\Sigma^2 v \le \sqrt{n}\sigma_1^2$, and take $\mu > \sqrt{n}\|Z(0)\|_2^2$. Since $\|Z(0)\|_2 \le \sqrt{m}\|Z(0)\|_1$, one can also use $\mu > m\sqrt{n}\|Z(0)\|_1^2$, where the bound is computationally less expensive to evaluate.

We point out that the ODE (28) takes the form of Oja's learning equation [Oja82, Oja89]

$$Z'(t) = (I - Z(t)Z^T(t))\, C\, Z(t), \tag{30}$$

where C depends on the state according to $C(Z) = \lfloor ZZ^T\mathbf{1}\rceil$. As such, (28) admits the interpretation of a "fake" gradient flow

$$Z'(t) = \mathrm{grad}\, f_{C(Z(t))}(Z(t)), \tag{31}$$

where $f_C : \mathbb{R}_*^{n\times m} \to \mathbb{R} : Z \mapsto \frac{1}{2}\mathrm{tr}(Z^T C Z)$ and the gradient is taken with respect to the Frobenius inner product; this follows, e.g., from [AMS08, §4.8.1]. (The gradient if "fake" because the argument Z also appears as a parameter: $\mathrm{grad}\, f_{C(Z)}(Z)$ differs from $\mathrm{grad}\,(Z \mapsto f_{C(Z)}(Z))(Z)$.) The flow (30) can also be thought of as a realization in $\mathrm{St}(m,n)$ of the gradient flow of $f_C : \mathrm{Gr}(m,n) \to \mathbb{R} : \mathsf{span}(Z) \mapsto \frac{1}{2}\mathrm{tr}((Z^T Z)^{-1} Z^T C Z)$; see, e.g., [AMS08, §4.9.1].

It would be worth investigating how much of the analysis of [YHM94], that concerns the case where C is constant, applies to $C(Z) = \lfloor ZZ^T\mathbf{1}\rceil$. The connection with Oja's flow also opens a way for applying the geometric Newton of [AIDV08], that handles the fact that the zeros of the right-hand side of (30) are not isolated due to a symmetry by the right-action of the orthogonal group.

6. Dynamics on the set of rank-m projectors in $\mathbb{R}^n$. There is a well-known one-to-one correspondence between the set $\mathrm{Gr}(m,n)$ of all m-planes in $\mathbb{R}^n$ and the set

$$\mathrm{Pj}(m,n) = \{M \in \mathbb{R}^{n\times n} : M^T = M,\ MM = M,\ \mathrm{rank}(M) = m\}$$

of all rank-m orthogonal projectors in $\mathbb{R}^n$. The correspondence associates to an m-plane the orthogonal projector onto the m-plane: $\mathsf{span}(Z)$ in $\mathrm{Gr}(m,n)$ is associated to $Z(Z^TZ)^{-1}Z^T$ in $\mathrm{Pj}(m,n)$. (It is thus common to use the same notation $\mathrm{Gr}(m,n)$ for both sets.)

The flow on $\mathrm{Gr}(m,n)$ induced by (25)—and by (28) on $\mathrm{St}(m,n)$—is given on $\mathrm{Pj}(m,n)$ by

$$\begin{aligned}
M' &= \big(Z(Z^TZ)^{-1}Z^T\big)' && (32)\\
&= \left(\lfloor M\mathbf{1}\rceil - \frac{1}{2}I\right) M - Z(Z^TZ)^{-1}Z^TZ'(Z^TZ)^{-1}Z^T + sym\\
&= \lfloor M\mathbf{1}\rceil M - \frac{1}{2}M - M(\lfloor M\mathbf{1}\rceil - \frac{1}{2}I)M + sym\\
&= \lfloor M\mathbf{1}\rceil M - M\lfloor M\mathbf{1}\rceil M + sym\\
&= \lfloor M\mathbf{1}\rceil M + M\lfloor M\mathbf{1}\rceil - 2M\lfloor M\mathbf{1}\rceil M. && (33)
\end{aligned}$$

This is the flow (4) derived in [SSZ91] and further analyzed by Zhao in [Zha08]. We have thus shown the following:

PROPOSITION 4. *Let $t \mapsto Z(t)$, resp. $t \mapsto M(t)$, denote the solution trajectories of* (25)*, resp.* (4)*, such that $Z(0) = Z_0 \in \mathbb{R}_*^{n\times m}$, resp. $M(0) = M_0 \in \mathrm{Pj}(m,n)$. If* $\mathsf{span}(Z_0) = \mathsf{span}(M_0)$*, then* $\mathsf{span}(Z(t)) = \mathsf{span}(M(t))$ *for all t.*

Since all projectors M are idempotent (i.e., $MM = M$), it follows that (33) admits the "double bracket" formulation (6), i.e.,

$$M' = [M, [M, \lfloor M\mathbf{1}\rceil]], \tag{34}$$

where $[A,B] = AB - BA$ denotes the matrix commutator.

As we pointed out in the introduction, the "classical" double bracket flow

$$M' = [M, [M, N]], \tag{35}$$

where N is a constant matrix, has been widely studied in the literature since [Bro91]. In particular, its numerical integration was investigated in [Ise02, Cas04]. The flow (34) is more challenging than (35) as its "N" matrix depends on $M(t)$. Nevertheless, it can be hoped that the knowledge accumulated on (35) can be extended to some degree to (34).

For example, the flow (34) in the space $\mathcal{S}_{sym}(n)$ of all $n \times n$ real symmetric matrices is isospectral, that is, the spectrum of M is constant along the trajectories. This is because (34) still fits in the framework of isospectral flows of the form $M' = [M, B(M)]$ where $B(M)$ is skew-symmetric (see, e.g., [CIZ97]). As an aside, since $\mathrm{Pj}(m, n)$ is the subset of $\mathcal{S}_{sym}(n)$ consisting of all matrices with eigenvalue 1 with multiplicity m and 0 with multiplicity $n - m$, we recover the property that $\mathrm{Pj}(m, n)$ is an invariant of (34).

Observe also that the right-hand side of (34) can be interpreted as the orthogonal projection of $\lfloor M\mathbf{1} \rceil \in T_M \mathcal{S}_{sym}(n)$ onto $T_M \mathrm{Pj}(m, n)$ with respect to the Frobenius inner product [HHT07]. Moreover, much as the classical double bracket flow (35) can be interpreted as the gradient flow of

$$f_N : \mathrm{Pj}(m, n) \to \mathbb{R} : M \mapsto \mathrm{tr}(NM)$$

(see [HHT07]), the time-varying double bracket flow (34) can be interpreted as the "fake" gradient system

$$M' = \mathrm{grad}\, f_{\lfloor M\mathbf{1} \rceil}(M). \tag{36}$$

We have thus recovered the gradient formulation (31), now in the $\mathrm{Pj}(m, n)$ representation of $\mathrm{Gr}(m, n)$. (Observe that the two cost functions represent the same cost function on $\mathrm{Gr}(m, n)$, except for a factor of 2 that compensates for a discrepancy by a factor of 2 between the Frobenius inner product in $\mathrm{St}(m, n)$ and the Frobenius inner product in $\mathrm{Pj}(m, n)$.)

7. Comparison of the three representations. Propositions 2, 3, and 4 give us three different representations of the same subspace flow. The first one is the flow (25) on the noncompact Stiefel manifold $\mathbb{R}_*^{n \times m}$ that induces a subspace flow through the surjection $\mathsf{span} : \mathbb{R}_*^{n \times m} \to \mathrm{Gr}(m, n)$. The second one is the flow (28) on the orthogonal Stiefel manifold $\mathrm{St}(m, n)$ that induces a subspace flow through the surjection $\mathsf{span} : \mathrm{St}(m, n) \to \mathrm{Gr}(m, n)$. The last one is the flow (34) on the set $\mathrm{Pj}(m, n)$ of all rank-m orthogonal projectors in $\mathbb{R}^n$ that induces a subspace flow through the bijection $\mathsf{span} : \mathrm{Pj}(m, n) \to \mathrm{Gr}(m, n)$. Underlying these different representations of the same subspace flow is the representation of $\mathrm{Gr}(m, n)$ as the quotients manifolds $\mathbb{R}_*^{n \times m}/\mathrm{GL}(m)$, $\mathrm{St}(m, n)/\mathrm{O}(m)$ and as the submanifold $\mathrm{Pj}(m, n)$ of $\mathbb{R}^{n \times n}$, see, e.g., [EAS98, AMS04, AMS08, HHT07] for details.

The analysis of the projector form (34) carried out by Zhao [Zha08] yields results for the two other forms. In particular, Zhao shows that for all solution $M(t)$ of (4) on $\mathrm{Pj}(m, n)$, the limit $M(\infty) = \lim_{t \to \infty} M(t)$ exists, and $M(\infty)\mathbf{1}$ is a vector of ones and zeros. Let $\mathcal{B}$ denote the index set corresponding to the ones and $\mathcal{N}$ the index set corresponding to the zeros. If $M(0) = Z_0(Z_0^T Z_0)^{-1} Z_0^T$ with $Z_0 = \Gamma(A, b, c)$ as in (23), then $(\mathcal{B}, \mathcal{N})$ is the optimal basis of the linear program (2)–(3) with coefficients (A, b, c). The next result follows.

PROPOSITION 5. *Let (A, b, c) define a strictly feasible linear program* (2)–(3) *with A of full row rank. Let $Z_0 = \Gamma(A, b, c)$ with Γ as in* (23)*. Let $t \mapsto Z(t)$ be the solution trajectory of* (25) *with $Z(0) = Z_0$. Then $\lim_{t\to\infty} Z(t)(Z(t)^T Z(t))^{-1} Z(t)^T \mathbf{1}$ is a vector of ones and zeros. The index set of the ones is $\mathcal{B}$ and the index set of the zeros is $\mathcal{N}$, where $(\mathcal{B}, \mathcal{N})$ is the optimal basis of the linear program defined by (A, b, c). Moreover, if $\tilde{Z}_0 \in \mathrm{St}(m, n)$ is such that $\mathsf{span}(\tilde{Z}_0) = \mathsf{span}(Z_0)$, then the solution trajectory $t \mapsto \tilde{Z}(t)$ of* (28) *is such that*

$$\lim_{t\to\infty} \tilde{Z}(t)\tilde{Z}(t)^T \mathbf{1}$$

is a vector of ones and zeros, with the ones at the indices in $\mathcal{B}$ and the zeros at the indices in $\mathcal{N}$.

An advantage of the formulations on the Stiefel manifolds $\mathbb{R}_*^{n\times m}$ and $\mathrm{St}(m, n)$ is that, when $n \gg m$, the formulation (25) involves only nm variables, instead of n^2 variables for (34). This makes (25) a more promising approach, computationally speaking, when the dual has much fewer variables than inequality constraints.

A disadvantage of the flow (25) on the noncompact Stiefel manifold $\mathbb{R}_*^{n\times m}$ is that some elements of the matrix grow unbounded as $t \to \infty$. This is remedied in the flow (28) on the orthogonal Stiefel manifold $\mathrm{St}(m, n)$.

8. Acknowledgements. This work was performed in part while the author was visiting the University of Cambridge, the University of Maryland, and Florida State University in April and May 2008. It benefited in particular from discussions with Kyle Gallivan, Arieh Iserles, André Tits, and Gongyun Zhao. Special thanks to André Tits and his optimization group for going over this topic in a group meeting.

REFERENCES

[AIDV08] P.-A. Absil, M. Ishteva, L. De Lathauwer, and S. Van Huffel. *A geometric Newton method for Oja's vector field*. Neural Comput., 2008. doi:10.1162/neco.2008.04-08-749.

[AMS04] P.-A. Absil, R. Mahony, and R. Sepulchre. *Riemannian geometry of Grassmann manifolds with a view on algorithmic computation*. Acta Appl. Math., 80:2(2004), pp. 199–220.

[AMS08] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, NJ, 2008.

[ASM08] P.-A. Absil, R. Sepulchre, and R. Mahony. *Continuous-time subspace flows related to the symmetric eigenproblem*. Pac. J. Optim., 4:2(2008), pp. 179–194.

[BBR92] Anthony M. Bloch, Roger W. Brockett, and Tudor S. Ratiu. *Completely integrable gradient flows*. Comm. Math. Phys., 147:1(1992), pp. 57–74.

[BI04] Anthony M. Bloch and Arieh Iserles. *On the optimality of double-bracket flows*. Int. J. Math. Math. Sci., 2004:61-64(2004), pp. 3301–3319.

[Bro91] R. W. Brockett. *Dynamical systems that sort lists, diagonalize matrices, and solve linear programming problems*. Linear Algebra Appl., 146(1991), pp. 79–91.

[Cas04] Fernando Casas. *Numerical integration methods for the double-bracket flow*. J. Comput. Appl. Math., 166:2(2004), pp. 477–495.

[CD91] Moody T. Chu and Kenneth R. Driessel. *Constructing symmetric nonnegative matrices with prescribed eigenvalues by differential equations*. SIAM J. Math. Anal., 22:5(1991),pp. 1372–1387.

[Chu95] Moody T. Chu. *Constructing a Hermitian matrix from its diagonal entries and eigenvalues*. SIAM J. Matrix Anal. Appl., 16:1(1995), pp. 207–217.

[CIZ97] Mari Paz Calvo, Arieh Iserles, and Antonella Zanna. *Numerical solution of isospectral flows*. Math. Comp., 66:220(1997), pp. 1461–1486.

[Deh95] Jeroen Dehaene. *Continuous-time matrix algorithms, systolic algorithms and adaptive neural networks*. PhD thesis, Katholieke Universiteit Leuven, Faculteit Toegepaste Wetenschappen, Departement elektrotechniek-ESAT, Kard. Mercierlaan 94, 3001 Leuven, Belgium, 1995. ftp://ftp.esat.kuleuven.ac.be/pub/SISTA/dehaene/phd/.

[EAS98] Alan Edelman, Tomás A. Arias, and Steven T. Smith. *The geometry of algorithms with orthogonality constraints*. SIAM J. Matrix Anal. Appl., 20:2(1998), pp. 303–353.

[FO01] R. Felipe and F. Ongay. *Super Brockett equations: a graded gradient integrable system*. Comm. Math. Phys., 220:1(2001), pp. 95–104.

[HHT07] Uwe Helmke, Knut Hüper, and Jochen Trumpf. *Newton's method on Grassmann manifolds*, September 2007. arXiv:0709.2205v2.

[HM94] Uwe Helmke and John B. Moore. *Optimization and Dynamical Systems*. Communications and Control Engineering Series. Springer-Verlag London Ltd., London, 1994. With a foreword by R. Brockett.

[Ise02] Arieh Iserles. *On the discretization of double-bracket flows*. Found. Comput. Math., 2:3(2002), pp. 305–329.

[MHM05] Jonathan H. Manton, Uwe Helmke, and Iven M. Y. Mareels. *A dual purpose principal and minor component flow*. Systems Control Lett., 54:8(2005), pp. 759–769.

[Oja82] E. Oja. *A simplified neuron model as a principal component analyzer*. J. Math. Biol., 15(1982), pp. 267–273.

[Oja89] Erkki Oja. *Neural networks, principal components, and subspaces*. Int. J. Neural Syst., 1(1989), pp. 61–68.

[SSZ91] G. Sonnevend, J. Stoer, and G. Zhao. *On the complexity of following the central path of linear programs by linear extrapolation. II*. Math. Programming, 52(3, Ser. B):527–553 (1992), 1991. Interior point methods for linear programming: theory and practice (Scheveningen, 1990).

[Tam04] Tin-Yau Tam. *Gradient flows and double bracket equations*. *Differential Geom. Appl.*, 20:2(2004), pp. 209–224.

[Wri97] S. J. Wright. *Primal-Dual Interior-Point Methods*. SIAM, Philadelphia, 1997.

[Ye97] Yinyu Ye. *Interior point algorithms*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons Inc., New York, 1997. Theory and analysis, A Wiley-Interscience Publication.

[YHM94] Wei-Yong Yan, Uwe Helmke, and John B. Moore. *Global analysis of Oja's flow for neural networks*. IEEE Trans. on Neural Networks, 5:5(1994), pp. 674–683.

[Zha08] Gongyun Zhao. *Representing the space of linear programs as the Grassmann manifold*. Math. Program., July 2008. doi:10.1007/s10107-008-0237-6.