13

Matrix calculus

Let us first establish the notation. This is important, because bad notation is a serious obstacle to elegant mathematics and coherent exposition, and it can be misleading. If $\boldsymbol{f}$ is an $m \times 1$ vector function of an $n \times 1$ vector $\boldsymbol{x}$, then the *derivative* (or *Jacobian matrix*) of $\boldsymbol{f}$ is the $m \times n$ matrix

$$\mathrm{D}\boldsymbol{f}(\boldsymbol{x}) := \frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}'}, \tag{13.1}$$

the elements of which are the partial derivatives $\partial f_i(\boldsymbol{x})/\partial x_j$, $i = 1, \dots, m$, $j = 1, \dots, n$. There is no controversy about this definition. It implies, inter alia, that when $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, then $\partial \boldsymbol{y}/\partial \boldsymbol{x}' = \boldsymbol{A}$ (when $\boldsymbol{A}$ is a matrix of constants). It also implies that for a scalar function $\varphi(\boldsymbol{x})$, the derivative $\partial \varphi(\boldsymbol{x})/\partial \boldsymbol{x}'$ is a *row* vector, not a column vector.

Now consider an $m \times p$ *matrix* function $\boldsymbol{F}$ of an $n \times q$ *matrix* of variables $\boldsymbol{X}$. Clearly, the derivative is a matrix containing all $mpnq$ partial derivatives. Also, (13.1) should be a special case of the more general definition. The most obvious and elegant definition is

$$\mathrm{D}\boldsymbol{F}(\boldsymbol{X}) := \frac{\partial \operatorname{vec} \boldsymbol{F}(\boldsymbol{X})}{\partial (\operatorname{vec} \boldsymbol{X})'}, \tag{13.2}$$

which is an $mp \times nq$ matrix. As a result, if $\boldsymbol{F}$ is a function of a scalar x ($n = q = 1$), then $\mathrm{D}\boldsymbol{F}(x) = \partial \operatorname{vec} \boldsymbol{F}(x)/\partial x$, an $mp \times 1$ column vector. If φ is a scalar function of a matrix $\boldsymbol{X}$ ($m = p = 1$), then $\mathrm{D}\varphi(\boldsymbol{X}) = \partial \varphi(\boldsymbol{X})/\partial(\operatorname{vec} \boldsymbol{X})'$, a $1 \times nq$ row vector. The choice of ordering in (13.2) is not arbitrary. For example, the derivative of the scalar function $\varphi(\boldsymbol{X}) = \operatorname{tr}(\boldsymbol{X})$ is not $\mathrm{D}\varphi(\boldsymbol{X}) = \boldsymbol{I}_n$, but $\mathrm{D}\varphi(\boldsymbol{X}) = (\operatorname{vec} \boldsymbol{I}_n)'$.

For practical rather than theoretical reasons, the treatment of matrix calculus is based on *differentials* rather than derivatives. An important advantage is the following. Let $\boldsymbol{f}(\boldsymbol{x})$ be an $m \times 1$ vector function of an $n \times 1$ vector $\boldsymbol{x}$. Then the derivative $\mathrm{D}\boldsymbol{f}(\boldsymbol{x})$ is an $m \times n$ matrix, but the differential $\mathrm{d}\boldsymbol{f}(\boldsymbol{x})$ remains an $m \times 1$ vector. The advantage is even larger for matrices: $\mathrm{d}\boldsymbol{F}(\boldsymbol{X})$ has the same dimension as $\boldsymbol{F}$, irrespective of the dimension of $\boldsymbol{X}$.

Unless specified otherwise, φ denotes a scalar function, $\boldsymbol{f}$ a vector function, and $\boldsymbol{F}$ a matrix function. Also, x denotes a scalar argument, $\boldsymbol{x}$ a vector argument, and $\boldsymbol{X}$ a matrix argument. For example, we write

$$\varphi(x) = x^2, \quad \varphi(\boldsymbol{x}) = \boldsymbol{a}'\boldsymbol{x}, \quad \varphi(\boldsymbol{X}) = \operatorname{tr} \boldsymbol{X}'\boldsymbol{X},$$
$$\boldsymbol{f}(x) = (x, x^2)', \quad \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}, \quad \boldsymbol{f}(\boldsymbol{X}) = \boldsymbol{X}\boldsymbol{a},$$
$$\boldsymbol{F}(x) = x^2\boldsymbol{I}_m, \quad \boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{x}\boldsymbol{x}', \quad \boldsymbol{F}(\boldsymbol{X}) = \boldsymbol{X}'.$$

There is a possibility of confusion between the $'$ sign for derivative and transpose. Thus, the vector $\boldsymbol{f}(\boldsymbol{x})'$ will denote the transpose of $\boldsymbol{f}(\boldsymbol{x})$, while $\boldsymbol{f}'(\boldsymbol{x})$ will denote its derivative, and the same for scalar and matrix functions. However, we try and avoid the use of the $'$ sign for derivatives of vector or matrix functions.

Note carefully that all functions and variables in this chapter are real; that is, we only consider real-valued functions φ, $\boldsymbol{f}$, and $\boldsymbol{F}$ defined on a subset of $\mathbb{R}$, $\mathbb{R}^n$, or $\mathbb{R}^{n\times q}$. Special care needs to be taken when differentiating complex functions or real functions of complex variables, and we will not deal with these problems in this chapter.

In the one-dimensional case, the equation

$$\lim_{u\to 0} \frac{\varphi(x+u) - \varphi(x)}{u} = \varphi'(x)$$

defines the derivative of φ at x. Rewriting the equation gives

$$\varphi(x+u) = \varphi(x) + \varphi'(x)u + r_x(u),$$

where the remainder term $r_x(u)$ satisfies $r_x(u)/u \to 0$ as $u \to 0$. We now define the (first) differential of φ at x (with increment u) as $\mathrm{d}\varphi(x; u) = \varphi'(x)u$. For example, for $\varphi(x) = x^2$, we obtain $\mathrm{d}\varphi(x; u) = 2xu$. In practice we write $\mathrm{d}x$ instead of u, so that $\mathrm{d}\varphi(x) = \varphi'(x)\,\mathrm{d}x$ and, in the case $\varphi(x) = x^2$, $\mathrm{d}\varphi(x) = 2x\,\mathrm{d}x$. The double use of the symbol "d" requires careful justification, which is not provided in this chapter.

In the vector case we have

$$\boldsymbol{f}(\boldsymbol{x}+\boldsymbol{u}) = \boldsymbol{f}(\boldsymbol{x}) + (\mathrm{D}\boldsymbol{f}(\boldsymbol{x}))\boldsymbol{u} + \boldsymbol{r}_{\boldsymbol{x}}(\boldsymbol{u})$$

and the (first) differential is defined as $\mathrm{d}\boldsymbol{f}(\boldsymbol{x}; \boldsymbol{u}) = (\mathrm{D}\boldsymbol{f}(\boldsymbol{x}))\boldsymbol{u}$. The matrix case is obtained from the vector case by writing $\boldsymbol{f} := \operatorname{vec} \boldsymbol{F}$ and $\boldsymbol{x} := \operatorname{vec} \boldsymbol{X}$.

We need three crucial results: two identification results and one invariance result. The first identification result shows that the first derivative can be obtained (identified) from the first differential. We have

$$\mathrm{d}\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \iff \mathrm{D}\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}(\boldsymbol{x}),$$

where $\boldsymbol{A}(\boldsymbol{x})$, as the notation indicates, will in general depend on $\boldsymbol{x}$. More generally,

$$\mathrm{d}\operatorname{vec} \boldsymbol{F}(\boldsymbol{X}) = \boldsymbol{A}(\boldsymbol{X})\,\mathrm{d}\operatorname{vec} \boldsymbol{X} \iff \mathrm{D}\boldsymbol{F}(\boldsymbol{X}) = \boldsymbol{A}(\boldsymbol{X}). \tag{13.3}$$

For example, when $\varphi(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ $(\boldsymbol{A} = \boldsymbol{A}')$, then $\mathrm{d}\varphi = 2\boldsymbol{x}'\boldsymbol{A}\,\mathrm{d}\boldsymbol{x}$. Hence, $\mathrm{D}\varphi(\boldsymbol{x}) = 2\boldsymbol{x}'\boldsymbol{A}$.

The second identification result shows that the second derivative can be obtained (identified) from the second differential. We have

$$\mathrm{d}^2\varphi(\boldsymbol{x}) = (\mathrm{d}\boldsymbol{x})'\boldsymbol{B}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \iff \mathrm{H}\varphi(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{B}(\boldsymbol{x}) + \boldsymbol{B}(\boldsymbol{x})'), \tag{13.4}$$

where $\mathrm{H}\varphi(\boldsymbol{x})$ denotes the *Hessian matrix* with typical element $\partial^2\varphi/\partial x_i\partial x_j$. Notice that we present (13.4) only for scalar functions. It is possible to extend the result to vector functions and matrix functions, but this is seldom required. For example, when $\varphi(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ ($\boldsymbol{A} = \boldsymbol{A}'$), then $\mathrm{d}\varphi = 2\boldsymbol{x}'\boldsymbol{A}\,\mathrm{d}\boldsymbol{x}$ and

$$\mathrm{d}^2\varphi = 2\,\mathrm{d}\left(\boldsymbol{x}'\boldsymbol{A}\,\mathrm{d}\boldsymbol{x}\right) = 2(\mathrm{d}\boldsymbol{x})'\boldsymbol{A}\,\mathrm{d}\boldsymbol{x} + 2\boldsymbol{x}'\boldsymbol{A}\,\mathrm{d}^2\boldsymbol{x} = 2(\mathrm{d}\boldsymbol{x})'\boldsymbol{A}\,\mathrm{d}\boldsymbol{x},$$

because $\mathrm{d}^2\boldsymbol{x} = \boldsymbol{0}$, since $\boldsymbol{x}$ (trivially) is a linear function of $\boldsymbol{x}$. Hence, $\mathrm{H}\varphi(\boldsymbol{x}) = 2\boldsymbol{A}$. In this case the matrix $\boldsymbol{B} = 2\boldsymbol{A}$ is symmetric but this need not be the case in general. The Hessian matrix, however, *must* be symmetric, so we have to make it symmetric, as in (13.4).

The invariance result is essentially the chain rule. The chain rule tells us that the derivative of a composite function $h(x) = g(f(x))$ is given by

$$\mathrm{D}h(x) = \mathrm{D}g(f(x))\,\mathrm{D}f(x).$$

The equivalent result for differentials is called *Cauchy's rule of invariance*, and states that

$$\mathrm{d}h(x;u) = \mathrm{d}g(f(x);\mathrm{d}f(x;u)).$$

This looks more complicated than it is. For example, when $\varphi(x) = \sin x^2$, we can take $g(y) = \sin y$ and $f(x) = x^2$, so that $\mathrm{D}\varphi(x) = (\cos x^2)(2x)$. The differential is

$$\mathrm{d}\varphi = (\cos x^2)\,\mathrm{d}x^2 = (\cos x^2)(2x\,\mathrm{d}x).$$

Cauchy's rule thus allows sequential determination of the differential.

Special care needs to be taken when dealing with the second differential and the Hessian matrix of composite functions. Cauchy's invariance result is not applicable here. For example, if $\varphi(y) = \sin y$, then $\mathrm{d}\varphi = (\cos y)\,\mathrm{d}y$ and

$$\mathrm{d}^2\varphi = \mathrm{d}((\cos y)\,\mathrm{d}y) = (\mathrm{d}\cos y)\,\mathrm{d}y + (\cos y)\,\mathrm{d}^2 y = -(\sin y)(\mathrm{d}y)^2,$$

because $\mathrm{d}^2 y = 0$. However, if we are now told that $y = x^2$, then it is still true, by Cauchy's invariance rule, that $\mathrm{d}\varphi = (\cos y)\,\mathrm{d}x^2 = 2x(\cos x^2)\,\mathrm{d}x$, but for the second differential we have $\mathrm{d}^2\varphi \neq -(\sin y)(\mathrm{d}x^2)^2 = -4x^2(\sin x^2)(\mathrm{d}x)^2$. The reason is that $\mathrm{d}^2 y$ is no longer zero. There exists a chain rule for Hessian matrices, but in practice the simplest and safest procedure is to go back to the first differential. Then,

$$\begin{aligned}\mathrm{d}^2\varphi &= \mathrm{d}((\cos y)\,\mathrm{d}y) = -(\sin y)(\mathrm{d}y)^2 + (\cos y)\,\mathrm{d}^2 y \\ &= -(\sin x^2)(\mathrm{d}x^2)^2 + (\cos x^2)\,\mathrm{d}^2 x^2 = -4(\sin x^2)(x\,\mathrm{d}x)^2 + 2(\cos x^2)(\mathrm{d}x)^2 \\ &= \left(-4x^2\sin x^2 + 2\cos x^2\right)(\mathrm{d}x)^2.\end{aligned}$$

This works in precisely the same way for vector and matrix functions.

A major use of matrix calculus is in problems of optimization. Suppose we wish to minimize a scalar function $\varphi(\boldsymbol{X})$. We compute

$$\mathrm{d}\varphi = \sum_{ij} \frac{\partial \varphi}{\partial x_{ij}} \,\mathrm{d}x_{ij} = \operatorname{tr} \boldsymbol{A}' \,\mathrm{d}\boldsymbol{X},$$

where $\boldsymbol{A}$ will in general depend on $\boldsymbol{X}$, unless the function is linear. The first-order condition is thus $\boldsymbol{A}(\boldsymbol{X}) = \mathbf{O}$. In order to verify that the solution is a (local or global) minimum, various conditions are available. We only mention that if $\mathrm{d}^2\varphi \geq 0$, then φ is convex, and hence φ has a global minimum at the point where $\mathrm{d}\varphi = 0$; and, if $\mathrm{d}^2\varphi > 0$ for all $\mathrm{d}\boldsymbol{X} \neq \mathbf{O}$, then φ is *strictly* convex, so that φ has a strict global minimum at $\mathrm{d}\varphi = 0$.

More difficult is constrained optimization. This usually takes the form of minimizing $\varphi(\boldsymbol{X})$ subject to a matrix constraint $\boldsymbol{G}(\boldsymbol{X}) = \mathbf{O}$. We then define the Lagrangian function

$$\psi(\boldsymbol{X}) = \varphi(\boldsymbol{X}) - \operatorname{tr} \boldsymbol{L}'\boldsymbol{G}(\boldsymbol{X}),$$

where $\boldsymbol{L}$ is a matrix of Lagrange multipliers. (If $\boldsymbol{G}(\boldsymbol{X})$ happens to be symmetric, we may take $\boldsymbol{L}$ symmetric too.) If ψ is (strictly) convex, then φ has a (strict) global minimum at the point where $\mathrm{d}\psi = 0$ under the constraint $\boldsymbol{G}(\boldsymbol{X}) = \mathbf{O}$. The simplest case where this occurs is when φ is (strictly) convex and all constraints are linear.

In the first seven sections of this chapter we practice the use of the first differential and the first derivative. First we practice with the use of differentials (Section 13.1), then we discuss simple scalar, vector, and matrix functions (Sections 13.2–13.4), and then some more interesting functions: the inverse (Section 13.5), the exponential and logarithmic function (Section 13.6), and the determinant (Section 13.7).

The next two sections contain two important applications of matrix calculus. First, the evaluation of Jacobians. If $\boldsymbol{Y}$ is a one-to-one function of $\boldsymbol{X}$, then $\boldsymbol{J} := \partial \operatorname{vec} \boldsymbol{Y} / \partial (\operatorname{vec} \boldsymbol{X})'$ is the Jacobian matrix of the transformation and the absolute value of $\det(\boldsymbol{J})$ is the *Jacobian*. In Section 13.8 we show how matrix calculus can be used to obtain Jacobians, also (and in particular) when the matrix argument is symmetric. A second application is sensitivity analysis. Here we typically ask how an estimator or predictor changes with respect to small changes in some of its components, for example, how the OLS estimator $\hat{\boldsymbol{\beta}} := (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ changes with ("is sensitive to") small perturbations in $\boldsymbol{X}$. In Section 13.9 several examples demonstrate this approach.

Up to this point we did not need the second differential and the Hessian matrix. These are developed in Section 13.10.

Two further applications of matrix calculus are presented in the final three sections. Our third application is (constrained) optimization, which we demonstrate with least-squares problems, best linear (and quadratic) unbiased estimation (Section 13.11), and some simple maximum likelihood cases (Section 13.12). Finally, we consider inequalities. Every inequality can be considered as an optimization problem, because showing that $\varphi(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x}$ in S is equivalent to showing that the minimum of $\varphi(\boldsymbol{x})$ over all $\boldsymbol{x}$ in S is equal to zero. Thus, matrix calculus can often be fruitfully applied in proving inequalities (and even equalities, see Exercise 13.69).

13.1 Basic properties of differentials

Exercise 13.1 (Sum rules of differential) Let α be a constant, $\boldsymbol{A}$ a matrix of constants, and let $\boldsymbol{F}$ and $\boldsymbol{G}$ be two matrix functions of the same order. Show that:
(a) $\mathrm{d}\boldsymbol{A} = \mathbf{O}$;
(b) $\mathrm{d}(\alpha\boldsymbol{F}) = \alpha\,\mathrm{d}\boldsymbol{F}$;
(c) $\mathrm{d}(\boldsymbol{F}+\boldsymbol{G}) = \mathrm{d}\boldsymbol{F} + \mathrm{d}\boldsymbol{G}$;
(d) $\mathrm{d}(\boldsymbol{F}-\boldsymbol{G}) = \mathrm{d}\boldsymbol{F} - \mathrm{d}\boldsymbol{G}$;
(e) $\mathrm{d}\,\mathrm{tr}\,\boldsymbol{F} = \mathrm{tr}(\mathrm{d}\boldsymbol{F})$ ($\boldsymbol{F}$ square).

Solution
(a) Let $\varphi(x) := \alpha$ be a constant scalar function. Then its derivative $\varphi'(x)$ is zero, and hence $\mathrm{d}\varphi = \varphi'(x)\,\mathrm{d}x = 0$. The same holds for the matrix function, because the differential of a matrix is a matrix of differentials.
(b) This follows from the scalar result that $\mathrm{d}(\alpha\varphi(x)) = \alpha\,\mathrm{d}\varphi(x)$.
(c) Let us formally prove the case of a scalar function of a vector. Let $\varphi(\boldsymbol{x}) := f(\boldsymbol{x}) + g(\boldsymbol{x})$. Then,

$$\begin{aligned}\mathrm{d}\varphi(\boldsymbol{x};\boldsymbol{u}) &= \sum_j u_j \mathrm{D}_j\varphi(\boldsymbol{x}) = \sum_j u_j\left(\mathrm{D}_j f(\boldsymbol{x}) + \mathrm{D}_j g(\boldsymbol{x})\right)\\ &= \sum_j u_j \mathrm{D}_j f(\boldsymbol{x}) + \sum_j u_j \mathrm{D}_j g(\boldsymbol{x}) = \mathrm{d}f(\boldsymbol{x};\boldsymbol{u}) + \mathrm{d}g(\boldsymbol{x};\boldsymbol{u}).\end{aligned}$$

The matrix case then follows immediately.
(d)–(e) These are proved similarly. Since the derivative of a sum is the sum of the derivatives (linearity), the same holds for differentials.

Exercise 13.2 (Permutations of linear operators) For any matrix function $\boldsymbol{F}$, show that:
(a) $\mathrm{d}(\boldsymbol{F}') = (\mathrm{d}\boldsymbol{F})'$;
(b) $\mathrm{d}(\mathrm{vec}\,\boldsymbol{F}) = \mathrm{vec}(\mathrm{d}\boldsymbol{F})$.

Solution
Both results follow from the fact that the differential of a vector (matrix) is the vector (matrix) of differentials.

Exercise 13.3 (Product rules of differential) For any two conformable matrix functions $\boldsymbol{F}$ and $\boldsymbol{G}$, show that:
(a) $\mathrm{d}(\boldsymbol{F}\boldsymbol{G}) = (\mathrm{d}\boldsymbol{F})\boldsymbol{G} + \boldsymbol{F}(\mathrm{d}\boldsymbol{G})$;
(b) $\mathrm{d}(\boldsymbol{F}\otimes\boldsymbol{G}) = (\mathrm{d}\boldsymbol{F})\otimes\boldsymbol{G} + \boldsymbol{F}\otimes(\mathrm{d}\boldsymbol{G})$.

Solution
(a) We have

$$\begin{aligned}(\mathrm{d}(\boldsymbol{F}\boldsymbol{G}))_{ij} &= \mathrm{d}(\boldsymbol{F}\boldsymbol{G})_{ij} = \mathrm{d}\sum_k f_{ik}g_{kj} = \sum_k \mathrm{d}(f_{ik}g_{kj}) \\ &= \sum_k ((\mathrm{d}f_{ik})g_{kj} + f_{ik}\,\mathrm{d}g_{kj}) = \sum_k (\mathrm{d}f_{ik})g_{kj} + \sum_k f_{ik}\,\mathrm{d}g_{kj} \\ &= ((\mathrm{d}\boldsymbol{F})\boldsymbol{G})_{ij} + (\boldsymbol{F}\,\mathrm{d}\boldsymbol{G})_{ij}.\end{aligned}$$

(b) For a typical element of $\boldsymbol{F} \otimes \boldsymbol{G}$, say $f_{ij}g_{st}$, we have

$$\mathrm{d}(f_{ij}g_{st}) = (\mathrm{d}f_{ij})g_{st} + f_{ij}\,\mathrm{d}g_{st},$$

and the result follows.

13.2 Scalar functions

Exercise 13.4 (Linear, quadratic, and bilinear forms, vectors) Let $\boldsymbol{a}$ be a vector of constants and $\boldsymbol{A}$ a matrix of constants. Obtain the differential $\mathrm{d}\varphi$ and the derivative $\mathrm{D}\varphi$ of the following scalar functions:
(a) $\varphi(\boldsymbol{x}) := \boldsymbol{a}'\boldsymbol{x}$;
(b) $\varphi(\boldsymbol{x}) := \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$;
(c) $\varphi(\boldsymbol{x}_1, \boldsymbol{x}_2) := \boldsymbol{x}_1'\boldsymbol{A}\boldsymbol{x}_2$, a *bilinear form* in $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$.

Solution
(a) From $\mathrm{d}\varphi = \boldsymbol{a}'\,\mathrm{d}\boldsymbol{x}$, it follows that $\mathrm{D}\varphi = \boldsymbol{a}'$.
(b) We have $\mathrm{d}\varphi = (\mathrm{d}\boldsymbol{x})'\boldsymbol{A}\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{A}\,\mathrm{d}\boldsymbol{x} = \boldsymbol{x}'(\boldsymbol{A} + \boldsymbol{A}')\,\mathrm{d}\boldsymbol{x}$, and hence $\mathrm{D}\varphi = \boldsymbol{x}'(\boldsymbol{A} + \boldsymbol{A}')$. In quadratic forms there is no loss in generality if we take the matrix to be symmetric. If $\boldsymbol{A}$ is symmetric, the derivative reduces to $\mathrm{D}\varphi = 2\boldsymbol{x}'\boldsymbol{A}$, which agrees with the scalar case $\varphi(x) := ax^2$ with derivative $\mathrm{D}\varphi = 2ax$. (In general, it is a good idea to check vector and matrix derivatives with the scalar case.) The reason why we present also the derivative for the general, nonsymmetric case is that it is sometimes unpractical to first rewrite the quadratic form in its symmetric version.
(c) Let $\boldsymbol{x} := (\boldsymbol{x}_1', \boldsymbol{x}_2')'$. Then,

$$\begin{aligned}\mathrm{d}\varphi &= (\mathrm{d}\boldsymbol{x}_1)'\boldsymbol{A}\boldsymbol{x}_2 + \boldsymbol{x}_1'\boldsymbol{A}\,\mathrm{d}\boldsymbol{x}_2 = \boldsymbol{x}_2'\boldsymbol{A}'\,\mathrm{d}\boldsymbol{x}_1 + \boldsymbol{x}_1'\boldsymbol{A}\,\mathrm{d}\boldsymbol{x}_2 \\ &= (\boldsymbol{x}_1', \boldsymbol{x}_2')\begin{pmatrix}\mathbf{O} & \boldsymbol{A} \\ \boldsymbol{A}' & \mathbf{O}\end{pmatrix}\begin{pmatrix}\mathrm{d}\boldsymbol{x}_1 \\ \mathrm{d}\boldsymbol{x}_2\end{pmatrix} = \boldsymbol{x}'\boldsymbol{C}\,\mathrm{d}\boldsymbol{x}, \quad \text{where } \boldsymbol{C} := \begin{pmatrix}\mathbf{O} & \boldsymbol{A} \\ \boldsymbol{A}' & \mathbf{O}\end{pmatrix},\end{aligned}$$

implying that

$$\mathrm{D}\varphi = \partial\varphi/\partial\boldsymbol{x}' = \boldsymbol{x}'\boldsymbol{C} = (\boldsymbol{x}_2'\boldsymbol{A}' : \boldsymbol{x}_1'\boldsymbol{A}).$$

Exercise 13.5 (On the unit sphere) If $\boldsymbol{x}'\boldsymbol{x} = 1$ on an open subset S in $\mathbb{R}^n$, show that $\boldsymbol{x}'\,\mathrm{d}\boldsymbol{x} = 0$ on S.

Solution

If $x'x = 1$ at x *and* in a neighborhood of x, then

$$0 = \mathrm{d}(x'x) = (\mathrm{d}x)'x + x'\,\mathrm{d}x = 2x'\,\mathrm{d}x,$$

and the result follows.

Exercise 13.6 (Bilinear and quadratic forms, matrices) Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be two vectors of constants. Find the differential and derivative of the following scalar functions:
(a) $\varphi(\boldsymbol{X}) = \boldsymbol{a}'\boldsymbol{X}\boldsymbol{b}$, a bilinear form in $\boldsymbol{a}$ and $\boldsymbol{b}$;
(b) $\varphi(\boldsymbol{X}) = \boldsymbol{a}'\boldsymbol{X}\boldsymbol{X}'\boldsymbol{a}$;
(c) $\varphi(\boldsymbol{X}) = \boldsymbol{a}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{a}$.

Solution

(a) The differential is simply $\mathrm{d}\varphi = \boldsymbol{a}'(\mathrm{d}\boldsymbol{X})\boldsymbol{b}$. To obtain the derivative we have to write $\mathrm{d}\varphi = (\operatorname{vec}\boldsymbol{A})'\,\mathrm{d}\operatorname{vec}\boldsymbol{X}$ for some matrix $\boldsymbol{A}$. Hence, we rewrite $\mathrm{d}\varphi$ as

$$\mathrm{d}\varphi = \boldsymbol{a}'(\mathrm{d}\boldsymbol{X})\boldsymbol{b} = (\boldsymbol{b}' \otimes \boldsymbol{a}')\,\mathrm{d}\operatorname{vec}\boldsymbol{X}$$

with derivative

$$\mathrm{D}\varphi(\boldsymbol{X}) = \frac{\partial\varphi}{\partial(\operatorname{vec}\boldsymbol{X})'} = (\boldsymbol{b} \otimes \boldsymbol{a})'.$$

(b) We have

$$\begin{aligned}\mathrm{d}\varphi &= \boldsymbol{a}'(\mathrm{d}\boldsymbol{X})\boldsymbol{X}'\boldsymbol{a} + \boldsymbol{a}'\boldsymbol{X}(\mathrm{d}\boldsymbol{X})'\boldsymbol{a} = 2\boldsymbol{a}'(\mathrm{d}\boldsymbol{X})\boldsymbol{X}'\boldsymbol{a}\\ &= 2(\boldsymbol{a}'\boldsymbol{X} \otimes \boldsymbol{a}')\,\mathrm{d}\operatorname{vec}\boldsymbol{X},\end{aligned}$$

so that

$$\mathrm{D}\varphi(\boldsymbol{X}) = \frac{\partial\varphi(\boldsymbol{X})}{\partial(\operatorname{vec}\boldsymbol{X})'} = 2(\boldsymbol{X}'\boldsymbol{a} \otimes \boldsymbol{a})'.$$

(c) Similarly,

$$\mathrm{d}\varphi = \boldsymbol{a}'(\mathrm{d}\boldsymbol{X})'\boldsymbol{X}\boldsymbol{a} + \boldsymbol{a}'\boldsymbol{X}'(\mathrm{d}\boldsymbol{X})\boldsymbol{a} = 2\boldsymbol{a}'\boldsymbol{X}'(\mathrm{d}\boldsymbol{X})\boldsymbol{a} = 2(\boldsymbol{a}' \otimes \boldsymbol{a}'\boldsymbol{X}')\,\mathrm{d}\operatorname{vec}\boldsymbol{X},$$

yielding $\mathrm{D}\varphi(\boldsymbol{X}) = 2(\boldsymbol{a} \otimes \boldsymbol{X}\boldsymbol{a})'$.

Exercise 13.7 (Differential and trace) For a scalar function φ with differential $\mathrm{d}\varphi = \operatorname{tr}(\boldsymbol{A}'\,\mathrm{d}\boldsymbol{X})$, show that $\mathrm{D}\varphi(\boldsymbol{X}) = (\operatorname{vec}\boldsymbol{A})'$.

Solution

This is a very useful property, and simple to prove:

$$\mathrm{d}\varphi = \operatorname{tr}\boldsymbol{A}'\,\mathrm{d}\boldsymbol{X} = (\operatorname{vec}\boldsymbol{A})'\,\mathrm{d}\operatorname{vec}\boldsymbol{X} \iff \mathrm{D}\varphi(\boldsymbol{X}) = (\operatorname{vec}\boldsymbol{A})'.$$

Exercise 13.8 (Trace of powers, 1) Use Exercise 13.7 to obtain the differential and derivative of:
(a) $\varphi(\boldsymbol{X}) := \operatorname{tr}\boldsymbol{X}$;

(b) $\varphi(\boldsymbol{X}) := \operatorname{tr} \boldsymbol{X}^2$;
(c) $\varphi(\boldsymbol{X}) := \operatorname{tr} \boldsymbol{X}^p$.

Solution
(a) First,

$$\mathrm{d}\varphi = \mathrm{d}(\operatorname{tr} \boldsymbol{X}) = \operatorname{tr}(\mathrm{d}\boldsymbol{X}) = \operatorname{tr}(\boldsymbol{I}\,\mathrm{d}\boldsymbol{X}) \implies \mathrm{D}\varphi = (\operatorname{vec} \boldsymbol{I})'.$$

(b) Next,

$$\mathrm{d}\varphi = \mathrm{d}\operatorname{tr} \boldsymbol{X}^2 = \operatorname{tr}(\mathrm{d}\boldsymbol{X})\boldsymbol{X} + \operatorname{tr} \boldsymbol{X}\,\mathrm{d}\boldsymbol{X} = 2\operatorname{tr} \boldsymbol{X}\,\mathrm{d}\boldsymbol{X} \implies \mathrm{D}\varphi = 2(\operatorname{vec} \boldsymbol{X}')'.$$

(Notice the transpose of $\boldsymbol{X}$. This corresponds to the rule in Exercise 13.7 and also to the fact that $\partial \operatorname{tr} \boldsymbol{X}^2/\partial x_{ij} = 2x_{ji}$.)
(c) Finally,

$$\begin{aligned}\mathrm{d}\varphi = \operatorname{tr} \boldsymbol{X}^p &= \operatorname{tr}(\mathrm{d}\boldsymbol{X})\boldsymbol{X}^{p-1} + \operatorname{tr} \boldsymbol{X}(\mathrm{d}\boldsymbol{X})\boldsymbol{X}^{p-2} + \cdots + \operatorname{tr} \boldsymbol{X}^{p-1}(\mathrm{d}\boldsymbol{X})\\ &= p\operatorname{tr} \boldsymbol{X}^{p-1}\,\mathrm{d}\boldsymbol{X},\end{aligned}$$

implying that

$$\mathrm{D}\varphi = p(\operatorname{vec}(\boldsymbol{X}')^{p-1})'.$$

Exercise 13.9 (Trace of powers, 2) Find the differential and derivative of:
(a) $\varphi(\boldsymbol{X}) := \operatorname{tr} \boldsymbol{X}'\boldsymbol{X}$;
(b) $\varphi(\boldsymbol{X}) := \operatorname{tr}(\boldsymbol{X}'\boldsymbol{X})^p$;
(c) $\varphi(\boldsymbol{X}) := \operatorname{tr}(\boldsymbol{X}\boldsymbol{X}')^p$.
(d) What is the difference between the derivatives in (c) and (b)?

Solution
(a) From

$$\mathrm{d}\varphi = \operatorname{tr}(\mathrm{d}\boldsymbol{X})'\boldsymbol{X} + \operatorname{tr} \boldsymbol{X}'\,\mathrm{d}\boldsymbol{X} = 2\operatorname{tr} \boldsymbol{X}'\,\mathrm{d}\boldsymbol{X}$$

it follows that $\mathrm{D}\varphi(\boldsymbol{X}) = 2(\operatorname{vec} \boldsymbol{X})'$.
(b) More generally,

$$\begin{aligned}\mathrm{d}\varphi &= \operatorname{tr}(\mathrm{d}(\boldsymbol{X}'\boldsymbol{X}))(\boldsymbol{X}'\boldsymbol{X})^{p-1} + \cdots + \operatorname{tr}(\boldsymbol{X}'\boldsymbol{X})^{p-1}\,\mathrm{d}(\boldsymbol{X}'\boldsymbol{X})\\ &= p\operatorname{tr}(\boldsymbol{X}'\boldsymbol{X})^{p-1}\,\mathrm{d}(\boldsymbol{X}'\boldsymbol{X}) = p\operatorname{tr}(\boldsymbol{X}'\boldsymbol{X})^{p-1}(\mathrm{d}\boldsymbol{X})'\boldsymbol{X} + p\operatorname{tr}(\boldsymbol{X}'\boldsymbol{X})^{p-1}\boldsymbol{X}'\,\mathrm{d}\boldsymbol{X}\\ &= 2p\operatorname{tr}(\boldsymbol{X}'\boldsymbol{X})^{p-1}\boldsymbol{X}'\,\mathrm{d}\boldsymbol{X},\end{aligned}$$

with derivative

$$\mathrm{D}\varphi(\boldsymbol{X}) = 2p(\operatorname{vec} \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{p-1})'.$$

(c) Similarly,

$$\mathrm{d}\varphi = p\operatorname{tr}(\boldsymbol{X}\boldsymbol{X}')^{p-1}(\mathrm{d}\boldsymbol{X}\boldsymbol{X}') = 2p\operatorname{tr} \boldsymbol{X}'(\boldsymbol{X}\boldsymbol{X}')^{p-1}\,\mathrm{d}\boldsymbol{X}$$

implies

$$\mathrm{D}\varphi(\boldsymbol{X}) = 2p(\operatorname{vec}(\boldsymbol{X}\boldsymbol{X}')^{p-1}\boldsymbol{X})'.$$

(d) There is no difference between the two derivatives, because

$$\operatorname{tr}(\boldsymbol{X}'\boldsymbol{X})^p = \operatorname{tr}(\boldsymbol{X}'\boldsymbol{X})\cdots(\boldsymbol{X}'\boldsymbol{X}) = \operatorname{tr}\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{X}')\cdots(\boldsymbol{X}\boldsymbol{X}')\boldsymbol{X} = \operatorname{tr}(\boldsymbol{X}\boldsymbol{X}')^p$$

and

$$\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{p-1} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})\cdots(\boldsymbol{X}'\boldsymbol{X}) = (\boldsymbol{X}\boldsymbol{X}')\cdots(\boldsymbol{X}\boldsymbol{X}')\boldsymbol{X} = (\boldsymbol{X}\boldsymbol{X}')^{p-1}\boldsymbol{X}.$$

Exercise 13.10 (Linear and quadratic matrix forms) Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two matrices of constants. Find the differential and derivative of:
(a) $\varphi(\boldsymbol{X}) := \operatorname{tr}\boldsymbol{A}\boldsymbol{X}$;
(b) $\varphi(\boldsymbol{X}) := \operatorname{tr}\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}'\boldsymbol{B}$;
(c) $\varphi(\boldsymbol{X}) := \operatorname{tr}\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}$.

Solution
(a) From $\mathrm{d}\varphi = \operatorname{tr}\boldsymbol{A}\,\mathrm{d}\boldsymbol{X}$, we find $\mathrm{D}\varphi(\boldsymbol{X}) = (\operatorname{vec}\boldsymbol{A}')'$, in accordance with Exercise 13.7.
(b) From

$$\begin{aligned}\mathrm{d}\varphi &= \operatorname{tr}(\mathrm{d}\boldsymbol{X})\boldsymbol{A}\boldsymbol{X}'\boldsymbol{B} + \operatorname{tr}\boldsymbol{X}\boldsymbol{A}(\mathrm{d}\boldsymbol{X})'\boldsymbol{B} = \operatorname{tr}\boldsymbol{A}\boldsymbol{X}'\boldsymbol{B}\,\mathrm{d}\boldsymbol{X} + \operatorname{tr}\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{B}'(\mathrm{d}\boldsymbol{X})\\ &= \operatorname{tr}(\boldsymbol{A}\boldsymbol{X}'\boldsymbol{B} + \boldsymbol{A}'\boldsymbol{X}'\boldsymbol{B}')\,\mathrm{d}\boldsymbol{X},\end{aligned}$$

we obtain

$$\mathrm{D}\varphi(\boldsymbol{X}) = (\operatorname{vec}(\boldsymbol{B}'\boldsymbol{X}\boldsymbol{A}' + \boldsymbol{B}\boldsymbol{X}\boldsymbol{A}))'.$$

(c) And,

$$\mathrm{d}\varphi = \operatorname{tr}(\mathrm{d}\boldsymbol{X})\boldsymbol{A}\boldsymbol{X}\boldsymbol{B} + \operatorname{tr}\boldsymbol{X}\boldsymbol{A}(\mathrm{d}\boldsymbol{X})\boldsymbol{B} = \operatorname{tr}(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B} + \boldsymbol{B}\boldsymbol{X}\boldsymbol{A})\,\mathrm{d}\boldsymbol{X}$$

yields

$$\mathrm{D}\varphi(\boldsymbol{X}) = \left(\operatorname{vec}(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B} + \boldsymbol{B}\boldsymbol{X}\boldsymbol{A})'\right)'.$$

Exercise 13.11 (Sum of squares) Let $\varphi(\boldsymbol{X})$ be defined as the sum of the squares of all elements in $\boldsymbol{X}$. Obtain $\mathrm{d}\varphi$ and $\mathrm{D}\varphi$.

Solution
The trick here is to work with the *matrix* $\boldsymbol{X}$ rather than with the individual *elements* of $\boldsymbol{X}$. Thus we write

$$\varphi(\boldsymbol{X}) = \sum_i \sum_j x_{ij}^2 = \operatorname{tr}\boldsymbol{X}'\boldsymbol{X}$$

and hence $\mathrm{d}\varphi = 2\operatorname{tr}\boldsymbol{X}'\,\mathrm{d}\boldsymbol{X}$, and $\mathrm{D}\varphi(\boldsymbol{X}) = 2(\operatorname{vec}\boldsymbol{X})'$.

Exercise 13.12 (A selector function) Let $\varphi(\boldsymbol{X})$ be defined as the ij-th element of $\boldsymbol{X}^2$. Obtain $\mathrm{d}\varphi$ and $\mathrm{D}\varphi$.

Solution

As in Exercise 13.11, we want to work with the matrix $\boldsymbol{X}$, rather than with its elements. Let $\boldsymbol{e}_i$ denote the i-th unit vector, having 1 in its i-th position and zeros elsewhere. Then, $\varphi(\boldsymbol{X}) = \boldsymbol{e}_i'\boldsymbol{X}^2\boldsymbol{e}_j$ and

$$\mathrm{d}\varphi = \boldsymbol{e}_i'(\mathrm{d}\boldsymbol{X})\boldsymbol{X}\boldsymbol{e}_j + \boldsymbol{e}_i'\boldsymbol{X}(\mathrm{d}\boldsymbol{X})\boldsymbol{e}_j = \operatorname{tr}(\boldsymbol{X}\boldsymbol{e}_j\boldsymbol{e}_i' + \boldsymbol{e}_j\boldsymbol{e}_i'\boldsymbol{X})\,\mathrm{d}\boldsymbol{X},$$

so that the derivative takes the form

$$\mathrm{D}\varphi(\boldsymbol{X}) = (\operatorname{vec}(\boldsymbol{e}_i(\boldsymbol{X}\boldsymbol{e}_j)' + (\boldsymbol{X}'\boldsymbol{e}_i)\boldsymbol{e}_j'))' = (\operatorname{vec}(\boldsymbol{e}_i\boldsymbol{x}_{\bullet j}' + \boldsymbol{x}_{i\bullet}\boldsymbol{e}_j'))'.$$

13.3 Vector functions

Exercise 13.13 (Vector functions of a vector, 1) Obtain the differential and derivative of the vector functions:
(a) $\boldsymbol{f}(\boldsymbol{x}) := \boldsymbol{A}\boldsymbol{x}$ ($\boldsymbol{A}$ constant);
(b) $\boldsymbol{f}(\boldsymbol{x}) := \boldsymbol{A}\boldsymbol{g}(\boldsymbol{x})$ ($\boldsymbol{A}$ constant).
(c) What happens in (a) if the elements of $\boldsymbol{A}$ also depend on $\boldsymbol{x}$?

Solution

(a) Since $\mathrm{d}\boldsymbol{f} = \boldsymbol{A}\,\mathrm{d}\boldsymbol{x}$, we have $\mathrm{D}\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}$.
(b) Now we have

$$\mathrm{d}\boldsymbol{f} = \boldsymbol{A}\,\mathrm{d}\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{A}(\mathrm{D}\boldsymbol{g}(\boldsymbol{x}))\,\mathrm{d}\boldsymbol{x},$$

so that

$$\mathrm{D}\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}\mathrm{D}\boldsymbol{g}(\boldsymbol{x}).$$

(c) If $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}(\boldsymbol{x})\boldsymbol{x}$, then

$$\begin{aligned}\mathrm{d}\boldsymbol{f} &= (\mathrm{d}\boldsymbol{A})\boldsymbol{x} + \boldsymbol{A}\,\mathrm{d}\boldsymbol{x} = (\boldsymbol{x}'\otimes\boldsymbol{I})\,\mathrm{d}\operatorname{vec}\boldsymbol{A} + \boldsymbol{A}\,\mathrm{d}\boldsymbol{x}\\ &= \left((\boldsymbol{x}'\otimes\boldsymbol{I})\frac{\partial\operatorname{vec}\boldsymbol{A}}{\partial\boldsymbol{x}'} + \boldsymbol{A}\right)\mathrm{d}\boldsymbol{x},\end{aligned}$$

implying that

$$\mathrm{D}\boldsymbol{f}(\boldsymbol{x}) = \frac{\partial\boldsymbol{f}(\boldsymbol{x})}{\partial\boldsymbol{x}'} = (\boldsymbol{x}'\otimes\boldsymbol{I})\frac{\partial\operatorname{vec}\boldsymbol{A}}{\partial\boldsymbol{x}'} + \boldsymbol{A}.$$

Exercise 13.14 (Vector functions of a vector, 2)
(a) Let $\boldsymbol{f}(\boldsymbol{x}) := (\boldsymbol{x}'\boldsymbol{x})\boldsymbol{a}$, where $\boldsymbol{a}$ is a vector of constants. Find the differential and derivative.
(b) What happens if $\boldsymbol{a}$ also depends on $\boldsymbol{x}$?

Solution

(a) From $\mathrm{d}\boldsymbol{f} = (2\boldsymbol{x}'\,\mathrm{d}\boldsymbol{x})\boldsymbol{a} = 2\boldsymbol{a}\boldsymbol{x}'\,\mathrm{d}\boldsymbol{x}$, we obtain $\mathrm{D}\boldsymbol{f}(\boldsymbol{x}) = 2\boldsymbol{a}\boldsymbol{x}'$.

(b) If $\boldsymbol{a} = \boldsymbol{a}(\boldsymbol{x})$, then

$$\begin{aligned}\mathrm{d}\boldsymbol{f} &= (\mathrm{d}\boldsymbol{x}'\boldsymbol{x})\boldsymbol{a} + (\boldsymbol{x}'\boldsymbol{x})\,\mathrm{d}\boldsymbol{a} = (2\boldsymbol{x}'\,\mathrm{d}\boldsymbol{x})\boldsymbol{a} + \boldsymbol{x}'\boldsymbol{x}(\mathrm{D}\boldsymbol{a}(\boldsymbol{x}))\,\mathrm{d}\boldsymbol{x} \\ &= (2\boldsymbol{a}\boldsymbol{x}' + \boldsymbol{x}'\boldsymbol{x}\mathrm{D}\boldsymbol{a}(\boldsymbol{x}))\,\mathrm{d}\boldsymbol{x}\end{aligned}$$

so that

$$\mathrm{D}\boldsymbol{f}(\boldsymbol{x}) = 2\boldsymbol{a}\boldsymbol{x}' + \boldsymbol{x}'\boldsymbol{x}\mathrm{D}\boldsymbol{a}(\boldsymbol{x}).$$

Exercise 13.15 (Vector functions of a matrix) Let $\boldsymbol{a}$ be a vector of constants. Find the differential and derivative of the vector functions:

(a) $\boldsymbol{f}(\boldsymbol{X}) := \boldsymbol{X}\boldsymbol{a}$;

(b) $\boldsymbol{f}(\boldsymbol{X}) := \boldsymbol{X}'\boldsymbol{a}$.

Solution

(a) We have

$$\mathrm{d}\boldsymbol{f} = (\mathrm{d}\boldsymbol{X})\boldsymbol{a} = (\boldsymbol{a}' \otimes \boldsymbol{I})\,\mathrm{d}\operatorname{vec}\boldsymbol{X},$$

and hence

$$\mathrm{D}\boldsymbol{f}(\boldsymbol{x}) = \frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial(\operatorname{vec}\boldsymbol{X})'} = \boldsymbol{a}' \otimes \boldsymbol{I}.$$

(b) Similarly,

$$\begin{aligned}\mathrm{d}\boldsymbol{f} &= (\mathrm{d}\boldsymbol{X})'\boldsymbol{a} = \operatorname{vec}\left((\mathrm{d}\boldsymbol{X})'\boldsymbol{a}\right) \\ &= \operatorname{vec}\left(\boldsymbol{a}'\,\mathrm{d}\boldsymbol{X}\right) = (\boldsymbol{I} \otimes \boldsymbol{a}')\,\mathrm{d}\operatorname{vec}\boldsymbol{X},\end{aligned}$$

so that

$$\mathrm{D}\boldsymbol{f}(\boldsymbol{X}) = \boldsymbol{I} \otimes \boldsymbol{a}'.$$

13.4 Matrix functions

Exercise 13.16 (Matrix function of a vector) Obtain the differential and derivative of $\boldsymbol{F}(\boldsymbol{x}) := \boldsymbol{x}\boldsymbol{x}'$.

Solution

Since $\mathrm{d}\boldsymbol{F} = (\mathrm{d}\boldsymbol{x})\boldsymbol{x}' + \boldsymbol{x}(\mathrm{d}\boldsymbol{x})'$, we find

$$\mathrm{d}\operatorname{vec}\boldsymbol{F} = (\boldsymbol{x} \otimes \boldsymbol{I})\,\mathrm{d}\operatorname{vec}\boldsymbol{x} + (\boldsymbol{I} \otimes \boldsymbol{x})\,\mathrm{d}\operatorname{vec}\boldsymbol{x}' = (\boldsymbol{x} \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \boldsymbol{x})\,\mathrm{d}\boldsymbol{x},$$

so that

$$\mathrm{D}\boldsymbol{F} = \frac{\partial \operatorname{vec}\boldsymbol{F}(\boldsymbol{x})}{\partial \boldsymbol{x}'} = \boldsymbol{x} \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \boldsymbol{x}.$$