On Representing Polynomials by Matrices

Geraint Jones
Wolfson College
Oxford OX2 6PS

Irina Voiculescu
Oxford University Computing Laboratory
Oxford OX1 3QD

July 14, 2000

Abstract

There is a natural representation of polynomials of up to two variables by matrices, in which matrix multiplication and transposition are the only operations that are needed to perform standard manipulations on polynomials. This note records a generalisation to more than two variables, which did not seem obvious to us until after we found it.

1 Polynomials as vectors

A polynomial of degree n in one unknown, x,

$$\sum_{0 \le i \le n} a_i x^i$$

can be seen as an element of a vector space of dimension $n+1$ for which the powers $x^0, x^1, \ldots x^n$ are an adequate basis. Such a polynomial is evaluated by taking the inner product $\mathbf{A} \cdot \mathbf{x}$ of the scalar coefficients with the basis,

$$\mathbf{x} = \begin{pmatrix} 1 & x & x^2 & \ldots & x^n \end{pmatrix}^{\mathsf{T}}$$

and the inner product can be thought of as the matrix multiplication $\mathbf{x}^{\mathsf{T}}\mathbf{A}$, or of course by $\mathbf{A}^{\mathsf{T}}\mathbf{x}$.

This generalises in a straightforward way to polynomials in two unknowns, for example

$$\sum_{\substack{0 \le i \le n \\ 0 \le j \le m}} a_{ij} x^i y^j = \mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{y}$$

The apparent asymmetry between $\mathbf{x}$ and $\mathbf{y}$ is illusory, since $\mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{y} = \mathbf{y}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}\mathbf{x}$.

2 Change of basis

Suppose that $\mathbf{p}_x$ is some alternative basis for polynomials of degree n over x, then there will be a non-singular matrix $\mathbf{P}_{(n)}$ for which $\mathbf{p}_x = \mathbf{P}_{(n)}\mathbf{x}$.

For example the Bernstein polynomials of degree n on $[0,1]$.

$$\begin{aligned} B_i^{(n)}(x) &= \binom{n}{i} x^i (1-x)^{n-i} \\ &= \sum_{0 \le j \le n-i} \binom{n}{i}\binom{n-i}{j} x^i (-x)^{n-i-j} \\ &= \sum_{i \le k \le n} \binom{n}{k}\binom{k}{i} (-1)^{k-i}\, x^k \\ &= \mathbf{B}_{(n)}\mathbf{x} \end{aligned}$$

are a basis for the space of polynomials of degree n, and

$$\mathbf{b}_x^{(n)} = \mathbf{B}_{(n)}\mathbf{x}$$

where $\mathbf{B}_{(n)}$ is a triangular matrix.

A general polynomial in two variables can be re-represented in a new basis by calculating

$$\begin{aligned} \mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{y} &= (\mathbf{P}_{(n)}^{-1}\mathbf{p}_x)^{\mathsf{T}}\mathbf{A}(\mathbf{P}_{(m)}^{-1}\mathbf{p}_y) \\ &= \mathbf{p}_x^{\mathsf{T}}((\mathbf{P}_{(n)}^{-1})^{\mathsf{T}}\mathbf{A}\mathbf{P}_{(m)}^{-1})\mathbf{p}_y \end{aligned}$$

The polynomial represented by $\mathbf{A}$ in the standard (power) basis is therefore represented by $(\mathbf{P}_{(n)}^{\mathsf{T}})^{-1}\mathbf{A}\mathbf{P}_{(m)}^{-1}$ in the new basis.

This shows how to implement any change of basis in which the two unknowns are treated separately: it does not deal with a transformation into a basis in which the unknowns become entangled with each other.

3 Polynomials in more than two unknowns

The coefficients of a polynomial in three or more unknowns are naturally thought of as a higher-dimensional matrix, for example

$$p(x,y,z) = \sum_{\substack{0\leq i\leq n\\0\leq j\leq m\\0\leq k\leq q}} a_{ijk}x^i y^j z^k$$

The contrivance of using both pre- and post-multiplication in the case of two unknowns does not generalise to three or more variables. It is natural to try to define a family of operators $\otimes_x, \otimes_y, \otimes_z$ such that

$$p(x,y,z) = ((\mathbf{A} \otimes_x \mathbf{x}) \otimes_y \mathbf{y}) \otimes_z \mathbf{z}$$

however care is necessary to preserve the symmetrical treatment of the unknowns. The property corresponding to the associativity of pre- and post-multiplication by a vector in the bivariate case is that it must be possible to re-order the $\otimes_i$ operations, so that for example

$$p(x,y,z) = ((\mathbf{A} \otimes_y \mathbf{y}) \otimes_z \mathbf{z}) \otimes_x \mathbf{x}$$

and in particular that this must be well-defined. That is, each $\mathbf{A} \otimes_i \mathbf{v}$ must itself be a three-dimensional matrix.

In the bivariate case there is a natural pun between vectors of size n and $n \times 1$ matrices, so that pre- and post-multiplication by a vector can be identified with matrix multiplication, and the result $\mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{y}$ is punned with two dimensional matrix with only one entry in it.

In order to make this pun explicit, for convenience later in the calculation, define the promotion of a vector to a two-dimensional matrix by

$$(\overline{\overline{\mathbf{v}}})_{i,0} = \mathbf{v}_i$$

and the product operators by

$$\begin{aligned}(\mathbf{A} \otimes_x \mathbf{m})_{i,j,k} &= \sum_l \mathbf{A}_{l,j,k}\mathbf{m}_{l,i}\\ (\mathbf{A} \otimes_y \mathbf{m})_{i,j,k} &= \sum_l \mathbf{A}_{i,l,k}\mathbf{m}_{l,j}\\ (\mathbf{A} \otimes_z \mathbf{m})_{i,j,k} &= \sum_l \mathbf{A}_{i,j,l}\mathbf{m}_{l,k}\end{aligned}$$

so that

$$p(x,y,z) = (((\mathbf{A} \otimes_x \overline{\overline{\mathbf{x}}}) \otimes_y \overline{\overline{\mathbf{y}}}) \otimes_z \overline{\overline{\mathbf{z}}})_{0,0,0}.$$

This generalises to the case of d unknowns in a natural way which nevertheless strains the notation.

The coefficient matrix will be indexed by an d-component index, I, which will always range over a cuboid bounded by zero and the opposite vertex. Write $I < N$ for $\forall j : 0 \leq j < d. I_j < N_j$, and $I[p \to q]$ for the index in which the pth component has been substituted by q:

$$I[p \to q]_j = \begin{cases} q & \text{if } p = j\\ I_j & \text{otherwise}\end{cases}$$

Then for each p define $\otimes_p$ by

$$(\mathbf{A} \otimes_p \mathbf{m})_I = \sum_{0\leq i<N_p} \mathbf{A}_{I[p\to i]}\mathbf{m}_{i,I_p}$$

so that

$$\begin{aligned}&p(x_0, x_1, x_2, \ldots)\\ &\quad= \quad (\ldots(((\mathbf{A} \otimes_0 \overline{\overline{\mathbf{x}_0}}) \otimes_1 \overline{\overline{\mathbf{x}_1}}) \otimes_2 \overline{\overline{\mathbf{x}_2}}) \ldots)_{0,0,0,\ldots}\end{aligned}$$

which we will write as

$$\mathbf{A} \bigotimes_{0\leq i<d} \overline{\overline{\mathbf{x}_i}}$$

All expressions in $\otimes_j$ are to be read as associating to the left.

In order to preserve symmetry, it is necessary that

$$\mathbf{A} \otimes_p \mathbf{p} \otimes_q \mathbf{q} = \mathbf{A} \otimes_q \mathbf{q} \otimes_p \mathbf{p}$$

whenever $p \neq q$, which is proved below.

Moreover, since $\mathbf{A} \otimes_p \mathbf{B} \otimes_p \mathbf{C} = \mathbf{A} \otimes_p (\mathbf{BC})$, if there is a change of basis $\mathbf{p}_{(j)} = \mathbf{P}_{(j)}\mathbf{v}$ then

$$\begin{aligned}\mathbf{A} \otimes_j \overline{\overline{\mathbf{v}}} &= \mathbf{A} \otimes_j \overline{\overline{(\mathbf{P}_{(j)}^{-1}\mathbf{p}_{(j)})}}\\ &= \mathbf{A} \otimes_j (\mathbf{P}_{(j)}^{-1}\overline{\overline{\mathbf{p}_{(j)}}})\\ &= (\mathbf{A} \otimes_j \mathbf{P}_{(j)}^{-1}) \otimes_j \overline{\overline{\mathbf{p}_{(j)}}}\end{aligned}$$

so that

$$\mathbf{A} \bigotimes_i \overline{\overline{\mathbf{v}_i}} = (\mathbf{A} \bigotimes_i \mathbf{P}_{(i)}^{-1}) \bigotimes_i \overline{\overline{\mathbf{p}_i}}$$

A Algebra of operations

Commutativity

Provided that $p \neq q$, for $0 \leq I < N$,

$$
\begin{aligned}
& (\mathbf{A} \otimes_p \mathbf{p} \otimes_q \mathbf{q})_I \\
= & \sum_{0 \leq i < N_q} (\mathbf{A} \otimes_p \mathbf{p})_{I[q \to i]} \mathbf{q}_{i, I_q} \\
= & \sum_{0 \leq i < N_q} \left(\sum_{0 \leq j < N_p} \mathbf{A}_{I[q \to i][p \to j]} \mathbf{p}_{j, I[q \to i]_p} \right) \mathbf{q}_{i, I_q} \\
= & \{ \text{ provided that } p \neq q \} \\
& \sum_{\substack{0 \leq i < N_q \\ 0 \leq j < N_p}} \mathbf{A}_{I[p \to j][q \to i]} \mathbf{p}_{j, I_p} \mathbf{q}_{i, I_q} \\
= & \sum_{0 \leq j < N_p} \left(\sum_{0 \leq i < N_q} \mathbf{A}_{I[p \to j][q \to i]} \mathbf{q}_{i, I[p \to j]_q} \right) \mathbf{p}_{j, I_p} \\
= & \sum_{0 \leq j < N_p} (\mathbf{A} \otimes_q \mathbf{q})_{I[p \to j]} \mathbf{p}_{j, I_p} \\
= & (\mathbf{A} \otimes_q \mathbf{q} \otimes_p \mathbf{p})_I
\end{aligned}
$$

Matrix product

$$
\begin{aligned}
& (\mathbf{A} \otimes_p (\mathbf{B}\mathbf{C}))_I \\
= & \sum_{0 \leq i < n} \mathbf{A}_{I[p \to i]} (\mathbf{B}\mathbf{C})_{i, I_p} \\
= & \{ \text{ provided } \mathbf{B} \text{ is } n \times m \text{ and } \mathbf{C} \text{ is } m \times N_p \} \\
& \sum_{0 \leq i < n} \mathbf{A}_{I[p \to i]} \sum_{0 \leq j < n} \mathbf{B}_{i,j} \mathbf{C}_{j, I_p} \\
= & \sum_{0 \leq j < m} \left(\sum_{0 \leq i < n} \mathbf{A}_{I[p \to i]} \mathbf{B}_{i,j} \right) \mathbf{C}_{j, I_p} \\
= & \sum_{0 \leq j < m} (\mathbf{A} \otimes_p \mathbf{B})_{I[p \to j]} \mathbf{C}_{j, I_p} \\
= & (\mathbf{A} \otimes_p \mathbf{B} \otimes_p \mathbf{C})_I
\end{aligned}
$$