Homework 1
Due: Sept 9, 2003

1. (5 pts) Simplify the following linear algebra expressions as far as possible (essentially, remove as many parenthases as you can). For each, describe the shape of each element as completely as you can (e.g., "if X is an $n \times n$ square matrix, then Y must be an $n \times m$ rectangular matrix, Z must be the same shape as X, P must be an m column vector, and Q can be an arbitrary rectangular matrix"). What is the shape of the resulting expression?

   (a) $(X+Y)^\mathsf{T}(X+Y)$

   $$\begin{aligned} &= (X^\mathsf{T}+Y^\mathsf{T})(X+Y) \\ &= X^\mathsf{T}(X+Y)+Y^\mathsf{T}(X+Y) \\ &= X^\mathsf{T}X + X^\mathsf{T}Y + Y^\mathsf{T}X + Y^\mathsf{T}Y \end{aligned}$$

   X and Y must be the same shape, but otherwise can be arbitrary rectangular or square matrices or vectors. If X and Y are both $n \times m$ then the result is an $m \times m$ square matrix. The special case of X and Y both column vectors is a scalar result.

   (b) $((OR)^{-1}(N^{-1}F^{-1}))^{-1}D$

   $$\begin{aligned} &= (N^{-1}F^{-1})^{-1}(OR)D \\ &= (FN)(OR)D \\ &= FNORD \end{aligned}$$

   The rule for matrix multiplication is that inner *dimensions between pairs of multiplicands must match. In this case, we have the additional constraint that the product OR must be invertible and, therefore, square, so that if O is $p \times q$, then R must be $q \times p$, etc. (Note that O and R need not be* individually *invertible, though the result must be of full rank, i.e., $p < q$.) Similary, F and N must* both *be square (because they must be individually invertible), $p \times p$. If D is $p \times s$, then the result of this expression (when it exists) is $p \times s$.*

   (c) $((A+B+C)^\mathsf{T}D(E+F-G-H)^\mathsf{T}I(J+K))^\mathsf{T}$

   $$= (J+K)^\mathsf{T}I^\mathsf{T}(E+F-G-H)D^\mathsf{T}(A+B+C)$$

   It would actually be acceptable to stop there — the main point of this excercise was to demonstrate some facility with transposes. But, since someone asked about expanding parens all the way, you can also chunk through that excercise:

$$
\begin{aligned}
=&(J^\mathsf{T} I^\mathsf{T} + K^\mathsf{T} I^\mathsf{T})(E + F - G - H)D^\mathsf{T}(A + B + C)\\
=&(J^\mathsf{T} I^\mathsf{T} E + J^\mathsf{T} I^\mathsf{T} F - J^\mathsf{T} I^\mathsf{T} G - J^\mathsf{T} I^\mathsf{T} H +\\
&K^\mathsf{T} I^\mathsf{T} E + K^\mathsf{T} I^\mathsf{T} F - K^\mathsf{T} I^\mathsf{T} G - K^\mathsf{T} I^\mathsf{T} H)D^\mathsf{T}(A + B + C)\\
=&(J^\mathsf{T} I^\mathsf{T} E D^\mathsf{T} + J^\mathsf{T} I^\mathsf{T} F D^\mathsf{T} - J^\mathsf{T} I^\mathsf{T} G D^\mathsf{T} - J^\mathsf{T} I^\mathsf{T} H D^\mathsf{T} +\\
&K^\mathsf{T} I^\mathsf{T} E D^\mathsf{T} + K^\mathsf{T} I^\mathsf{T} F D^\mathsf{T} - K^\mathsf{T} I^\mathsf{T} G D^\mathsf{T} - K^\mathsf{T} I^\mathsf{T} H D^\mathsf{T})(A + B + C)\\
=&J^\mathsf{T} I^\mathsf{T} E D^\mathsf{T} A + J^\mathsf{T} I^\mathsf{T} F D^\mathsf{T} A - J^\mathsf{T} I^\mathsf{T} G D^\mathsf{T} A - J^\mathsf{T} I^\mathsf{T} H D^\mathsf{T} A +\\
&K^\mathsf{T} I^\mathsf{T} E D^\mathsf{T} A + K^\mathsf{T} I^\mathsf{T} F D^\mathsf{T} A - K^\mathsf{T} I^\mathsf{T} G D^\mathsf{T} A - K^\mathsf{T} I^\mathsf{T} H D^\mathsf{T} A +\\
&J^\mathsf{T} I^\mathsf{T} E D^\mathsf{T} B + J^\mathsf{T} I^\mathsf{T} F D^\mathsf{T} B - J^\mathsf{T} I^\mathsf{T} G D^\mathsf{T} B - J^\mathsf{T} I^\mathsf{T} H D^\mathsf{T} B +\\
&K^\mathsf{T} I^\mathsf{T} E D^\mathsf{T} B + K^\mathsf{T} I^\mathsf{T} F D^\mathsf{T} B - K^\mathsf{T} I^\mathsf{T} G D^\mathsf{T} B - K^\mathsf{T} I^\mathsf{T} H D^\mathsf{T} B +\\
&J^\mathsf{T} I^\mathsf{T} E D^\mathsf{T} C + J^\mathsf{T} I^\mathsf{T} F D^\mathsf{T} C - J^\mathsf{T} I^\mathsf{T} G D^\mathsf{T} C - J^\mathsf{T} I^\mathsf{T} H D^\mathsf{T} C +\\
&K^\mathsf{T} I^\mathsf{T} E D^\mathsf{T} C + K^\mathsf{T} I^\mathsf{T} F D^\mathsf{T} C - K^\mathsf{T} I^\mathsf{T} G D^\mathsf{T} C - K^\mathsf{T} I^\mathsf{T} H D^\mathsf{T} C
\end{aligned}
$$

A, B, and C must all have the same shape ($n \times m$), D must be $n \times p$, E–H are all the same shape ($q \times p$), I is $q \times r$, and J and K are both $r \times s$. The full expression is $s \times m$.

A number of people pointed out that I usually denotes the identity matrix and (if we take $q == r$) can be dropped from the solution. They are, of course, quite correct, although I didn't think about this while I was writing the original problem...

(d) $(X + Y)^\mathsf{T}(Q + Z)(Q + Z)^\mathsf{T}(X + Y)$ where all elements are column vectors.

$$
\begin{aligned}
=&(X^\mathsf{T} + Y^\mathsf{T})(Q + Z)(Q^\mathsf{T} + Z^\mathsf{T})(X + Y)\\
=&(X^\mathsf{T} Q + Y^\mathsf{T} Q + X^\mathsf{T} Z + Y^\mathsf{T} Z)(Q^\mathsf{T} X + Q^\mathsf{T} Y + Z^\mathsf{T} X + Z^\mathsf{T} Y\\
=&X^\mathsf{T} Q Q^\mathsf{T} X + X^\mathsf{T} Q Q^\mathsf{T} Y + X^\mathsf{T} Q Z^\mathsf{T} X + X^\mathsf{T} Q Z^\mathsf{T} Y +\\
&Y^\mathsf{T} Q Q^\mathsf{T} X + Y^\mathsf{T} Q Q^\mathsf{T} Y + Y^\mathsf{T} Q Z^\mathsf{T} X + Y^\mathsf{T} Q Z^\mathsf{T} Y +\\
&X^\mathsf{T} Z Q^\mathsf{T} X + X^\mathsf{T} Z Q^\mathsf{T} Y + X^\mathsf{T} Z Z^\mathsf{T} X + X^\mathsf{T} Z Z^\mathsf{T} Y +\\
&Y^\mathsf{T} Z Q^\mathsf{T} X + Y^\mathsf{T} Z Q^\mathsf{T} Y + Y^\mathsf{T} Z Z^\mathsf{T} X + Y^\mathsf{T} Z Z^\mathsf{T} Y +
\end{aligned}
$$

The result is a scalar.

Alternatively, note that because all components are vectors, we have that $(X + Y)^\mathsf{T}(Q + Z)$ is a scalar, so we can write the original expression simply as $((X + Y)^\mathsf{T}(Q + Z))^2$.

(e) $(X + Y)^\mathsf{T}(Q + Z)(Q + Z)^\mathsf{T}(X + Y)$ where all elements are arbitrary rectangular matrices.

The form is the same as above, but the result is a rectangular matrix. If X and Y are $n \times m$ and Q and Z are $n \times r$, then the result is $m \times m$, a symmetric square matrix (and positive definite if all of the components are real and non-singular).

2. (5 pts) Let $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{b}$ be column vectors of length n, A be an $n \times n$ square matrix, and c be a scalar constant. Consider the function defined by:

$$f(\mathbf{x}) = \mathbf{x}^{\mathsf{T}} A \mathbf{x} + \mathbf{x}^{\mathsf{T}} \mathbf{y} \mathbf{y}^{\mathsf{T}} \mathbf{x} + \mathbf{x}^{\mathsf{T}} \mathbf{b} + c$$

(a) What is the shape of $f(\mathbf{x})$? (I.e., the shape of the result of the function.)

I meant to ask "shape, as in matrix, scalar, vector, etc.", in which case we have that $f(\mathbf{x})$ is a scalar. If you interpret "shape" to be a description of the form of the function in d-space, then this is a quadratic function — e.g., hyper-parabolic or somesuch.

(b) How many stationary points does this function have? Find an expression that describes the stationary point(s).

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{x}} f(\mathbf{x}) &= (A^{\mathsf{T}} + A)\mathbf{x} + ((yy^{\mathsf{T}})^{\mathsf{T}} + yy^{\mathsf{T}})\mathbf{x} + \mathbf{b} \\
&= (A^{\mathsf{T}} + A)\mathbf{x} + 2yy^{\mathsf{T}}\mathbf{x} + \mathbf{b} \\
&= (A^{\mathsf{T}} + A + 2yy^{\mathsf{T}})\mathbf{x} + \mathbf{b} \\
&= 0 \\
&\Rightarrow \\
\mathbf{x} &= -(A^{\mathsf{T}} + A + 2yy^{\mathsf{T}})^{-1}\mathbf{b}
\end{aligned}$$

(Assuming that you take $\frac{\partial}{\partial \mathbf{x}}$ to be a column *vector. If you would rather it be a* row *vector, then you have to transpose every term. It's sort-of a matter of taste.) The step from line 1 to 2 takes into account the fact that yy^{T} must be symmetric. In addition, if we happen to know that A is also symmetric, then we can reduce the last line to:*

$$\mathbf{x} = -\frac{1}{2}(A + yy^{\mathsf{T}})^{-1}\mathbf{b}$$

This system has either no solution (if $(A^{\mathsf{T}} + A + 2yy^{\mathsf{T}})$ is singular and the system is inconsistent) or one solution (if it's nonsingular) or an infinite number (if it's singular and the system is consistent), so the system posesses either no stationary points or one or ∞. We're usually concerned with the case of one solution, though, and mostly the systems we look at have a unique extremum.

(c) Suppose I tell you that a particular stationary point of this function is a maximum. What can you infer about the constituents of the function? What if it's a minimum?

If the stationary point is a maximum then the Hessian must be negative definite (and positive definite if the point is a minimum). It's not hard to show that the Hessian of this system is:

$$H(f(\mathbf{x})) = \frac{\partial}{\partial \mathbf{x}} \frac{\partial}{\partial \mathbf{x}} f(\mathbf{x}) = \frac{\partial^2}{\partial \mathbf{x}^2} = (A^{\mathsf{T}} + A + 2yy^{\mathsf{T}})$$

3. (5 pts) Consider the following system of equations

$$V_i = R_i + \gamma \sum_{j=1}^{n} T_{ij} V_j$$

where i and j both range $1..n$ and all quantities (V_i, R_i, etc.) are scalar.

(a) Write this system in linear algebraic notation (i.e., in terms of vectors and matrices rather than scalars). What "shape" is each – i.e., list which are vectors, which matrices, which scalars.

$$V = R + \gamma T V$$

Where V and R are column vectors with n components and T is a $n \times n$ matrix. γ is a scalar.

(b) Using your answer for 3a, solve for V in terms of the other quantities.

$$\begin{aligned} V &= R + \gamma T V \\ V - \gamma T V &= R \\ (I - \gamma T)V &= R \\ V &= (I - \gamma T)^{-1} R \end{aligned}$$

4. (5 pts) Given the following distribution over three discrete random variables X, Y, and Z, each taking on the values 0 or 1:

$X = 0$

	$Z = 0$	$Z = 1$
$Y = 0$	0	$\frac{1}{6}$
$Y = 1$	$\frac{1}{15}$	$\frac{1}{10}$

$X = 1$

	$Z = 0$	$Z = 1$
$Y = 0$	0	$\frac{1}{3}$
$Y = 1$	$\frac{2}{15}$	$\frac{1}{5}$

Find:

(a) The joint distributions of X and Y, X and Z, and Y and Z.

$P(X, Y) =$

	$X = 0$	$X = 1$
$Y = 0$	$1/6$	$1/3$
$Y = 1$	$1/6$	$1/3$

$P(X, Z) =$

	$X = 0$	$X = 1$
$Z = 0$	$1/15$	$2/15$
$Z = 1$	$4/15$	$8/15$

$P(Y, Z) =$

	$Z = 0$	$Z = 1$
$Y = 0$	0	$1/2$
$Y = 1$	$1/5$	$3/10$

(b) The marginal distributions of X, Y, and Z.

$$\begin{aligned} \Pr[X] &= [1/3 \;\; 2/3] \\ \Pr[Y] &= [1/2 \;\; 1/2] \\ \Pr[Z] &= [1/5 \;\; 4/5] \end{aligned}$$

(c) The conditional distribution of X given Y, X given Z, and Y given Z.

$P(X|Y) =$

	$Y = 0$	$Y = 1$
$X = 0$	$1/3$	$1/3$
$X = 1$	$2/3$	$2/3$

$P(X|Z) =$

	$Z = 0$	$Z = 1$
$X = 0$	$1/3$	$1/3$
$X = 1$	$2/3$	$2/3$

$P(Y|Z) =$

	$Z = 0$	$Z = 1$
$Y = 0$	0	$5/8$
$Y = 1$	1	$3/8$

Note that each column sums to 1. Also, note that both columns in each of the first two tables are identical (and are the same as the marginal of X). This gives you the answer to the next question:

(d) For each pair of RVs in this set, explain whether those two variables are statistically independent (including why).

X is statistically independent from both Y and Z. You can see this either because $P(X, Y) = P(X)P(Y)$ or because $P(X) = P(X|Y \in \{0, 1\})$ and similarly for Z. Y and Z are not *independent, however, because $P(Y, Z) \neq P(Y)P(Z)$ (in particular, $P(Y = 0, Z = 0) = 0$ but $P(Y = 0)P(Z = 0) = 1/10$).*

5. (10 pts) The joint distribution of a simple two-dimensional Gaussian random variable is given by the PDF:

$$f_{XY}(x, y) = \frac{1}{2\pi\sqrt{5}} e^{-\frac{1}{2}\left(\frac{1}{5}x^2 + y^2\right)}$$

(a) Describe the general structure of this PDF (e.g., what shape are the isopotential surfaces in 2-d space?)

The isopotential surfaces are defined to be the locus of (x, y) coordinates for which probability is constant. Examining the form of the Gaussian reveals that requiring $f(x, y) = c$ is equivalent to requiring that $\frac{1}{5}x^2 + y^2 = k$ for some k. It should be clear that this form is an ellipse centered on the origin with major axis along the x axis and minor axis in the y direction. The major axis is $\sqrt{5}$ times as long as the minor axis.

(b) Find $f_X(x)$ and $f_Y(y)$ (i.e., the marginal distributions of x and y). Describe each random variable.

The simplest solution to this part is to note that the PDF can be factored into separate components that depend only on x and on y and, therefore, the marginalization can be factored:

$$
\begin{aligned}
f(x,y) &= \frac{1}{2\pi\sqrt{5}} e^{-\frac{1}{2}\left(\frac{1}{5}x^2+y^2\right)} \\
&= \frac{1}{\sqrt{2\pi}\sqrt{5}} e^{-\frac{1}{2}\left(\frac{1}{5}x^2\right)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \\
&\Rightarrow \\
f(x) &= \int_{-\infty}^{\infty} f(x,y)dy \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sqrt{5}} e^{-\frac{1}{2}\left(\frac{1}{5}x^2\right)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy \\
&= \frac{1}{\sqrt{2\pi}\sqrt{5}} e^{-\frac{1}{2}\left(\frac{1}{5}x^2\right)} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy \\
&= \frac{1}{\sqrt{2\pi}\sqrt{5}} e^{-\frac{1}{2}\left(\frac{1}{5}x^2\right)}
\end{aligned}
$$

(Because the integral on the left is simply a Gaussian integrated over its whole range, which is just 1). The marginalization of y proceeds similarly and yields:

$$f(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$$

(c) Are x and y statistically independent? Why or why not?

Yes, because the joint can be written as a product of the marginals.

(d) Find $f_{X|Y}(x|y=3)$ and $f_{Y|X}(y|x=-1)$ (i.e., the conditional distributions of x given a fixed value of y and vice-versa).

By independence, the conditionals must be equal to the marginals, so $f_{X|Y}(x|y=3) = f_X(x)$ (from above). Similarly, $f_{Y|X}(y|x=-1) = f_Y(y)$ You can also see this by expanding $f_{XY}(x, y=3)/f_Y(y=3)$, etc.

6. (10 pts) The joint distribution of a more complicated two-dimensional Gaussian random variable is given by the PDF:

$$f_{XY}(x,y) = \frac{1}{2\pi\sqrt{5}} e^{-\frac{1}{2}\left(\frac{2}{5}x^2+\frac{4}{5}y^2+\frac{6}{5\sqrt{3}}xy+\left(\frac{4\sqrt{3}+12}{5\sqrt{3}}\right)x+\left(\frac{8\sqrt{3}-6}{5\sqrt{3}}\right)y+\frac{18\sqrt{3}-12}{5\sqrt{3}}\right)}$$

(a) Find the marginal distributions of x and y and describe each random variable.

This is somewhat algebraically painful, but not practically different from what you did in the previous question. There are two basic ways to go about this — the special case and the general.

In the special case, we tackle the integration directly. Consider, first, finding $f(x)$, i.e., the marginal of x. For the moment, we'll focus only on the exponent:

$$
-\frac{1}{2}\left(\frac{2}{5}x^2+\frac{4}{5}y^2+\frac{6}{5\sqrt{3}}xy+\left(\frac{-4\sqrt{3}+12}{5\sqrt{3}}\right)x+\left(\frac{16\sqrt{3}-6}{5\sqrt{3}}\right)y+\frac{18\sqrt{3}-12}{5\sqrt{3}}\right)
$$
$$
=-\frac{1}{10}\left(2x^2+4y^2+2\sqrt{3}xy+(-4+4\sqrt{3})x+(16-2\sqrt{3})y+18-4\sqrt{3}\right)
$$
$$
=-\frac{1}{10}\left(4y^2+(2\sqrt{3}x+16-2\sqrt{3})y+2x^2+(-4+4\sqrt{3})x+18-4\sqrt{3}\right)
$$

Now we complete the square on y. The rule for completing the square of ax^2+bx is $ax^2+bx+b^2/4a-b^2/4a=(\sqrt{a}x+b/2\sqrt{a})^2-b^2/4a$:

$$
\begin{aligned}
&=-\frac{1}{10}\left[4y^2+(2\sqrt{3}x+8-2\sqrt{3})y+\frac{(2\sqrt{3}x+16-2\sqrt{3})^2}{16}\right]+\\
&\quad-\frac{1}{10}\left[2x^2+(-4+4\sqrt{3})x+18-4\sqrt{3}-\frac{(2\sqrt{3}x+16-2\sqrt{3})^2}{16}\right]\\
&=-\frac{1}{10}\left(2y+\frac{(2\sqrt{3}x+16-2\sqrt{3})}{4}\right)^2+\\
&\quad-\frac{1}{10}\left[2x^2+(-4+4\sqrt{3})x+18-4\sqrt{3}-\frac{(2\sqrt{3}x+16-2\sqrt{3})^2}{16}\right]\\
&=-\frac{1}{2}\frac{4}{5}\left(y+\frac{(\sqrt{3}x+4-\sqrt{3})}{4}\right)^2+\\
&\quad-\frac{1}{10}h(x)
\end{aligned}
$$

So now we have separated the ugly thing into a pure function of x and a quadratic form of y. The original marginalization can now be rewritten:

$$
f(x)=\frac{1}{2\sqrt{2\pi}}e^{-\frac{1}{10}h(x)}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}(\sqrt{5/4})}e^{-\frac{1}{2}\frac{4}{5}\left(y+\frac{(\sqrt{3}x+4-\sqrt{3})}{4}\right)^2}dy
$$

At this point, it's obvious that the quantity inside the integral is, itself, a pure Gaussian in y, with variance $5/4$ and mean $\frac{(\sqrt{3}x+4-\sqrt{3})}{4}$. Because x is constant with respect to the integral, the whole quantity inside the integral reduces to 1. It's not immediately obvious, but the term $h(x)$ can be written,

$$
\begin{aligned}
h(x)&=\frac{5(x-1)^2}{4}\\
&\Rightarrow\\
f(x)&=\frac{1}{2\sqrt{2\pi}}e^{-\frac{1}{2}\frac{(x-1)^2}{4}}
\end{aligned}
$$

which is clearly a Gaussian with mean 1 *and std dev 2. A similar set of manipulations yields*

$$f(y) = \frac{1}{2\sqrt{\pi}} e^{-\frac{1}{2}\frac{(y+2)^2}{2}}$$

which is another Gaussian.

An alternate solution is to observe that the entire system is can be written in the linear algebraic form:

$$f(x,y) = \frac{1}{2\pi\sqrt{5}} e^{-\frac{1}{2}([x \;\; y]-[\mu_x \;\; \mu_y]) \left[\begin{array}{cc} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{array} \right]^{-1} \left(\left[\begin{array}{c} x \\ y \end{array} \right] - \left[\begin{array}{c} \mu_x \\ \mu_y \end{array} \right] \right)}$$

Expanding the exponent, you can match terms with the given exponent and use the fact that $|\Sigma|^{1/2} = \sqrt{5}$ *to determine that*

$$\left[\begin{array}{c} \mu_x \\ \mu_y \end{array} \right] = \left[\begin{array}{c} 1 \\ -2 \end{array} \right]$$
$$\Sigma = \left[\begin{array}{cc} 4 & -\sqrt{3} \\ -\sqrt{3} & 2 \end{array} \right]$$

from which the above marginals can be found. However *it was not adequate to merely show the resulting mean vector and covariance matrix and* state without proof *that the marginals are of the form given above. With no other knowledge, it's not obvious that the marginals actually turn out to be Gaussians themselves. This happens to be a theorem, but I wanted you either to show that theorem in detail (i.e., give a proof), or to carry out the integration to demonstrate that you do, in fact, reach Gaussian marginals in this case.*

(b) Are these variables statistically independent? Why or why not?

They are not, because $f(x,y) \neq f(x)f(y)$. *This turns out to be equivalent to observing that the coefficient of* xy *in the original exponent is non-zero, and is a consequence of* $\sigma_{xy} \neq 0$. *The two components are correlated and, therefore, non-independent.*