Spring 2015
Chris Christensen
MAT/CSC 483

Polygraphic Substitution Ciphers: The Hill Cipher, II

The Hill cipher

The Hill cipher was introduced in 1929 by Hunter College mathematician Lester Hill. Some things to recognize about the Hill cipher:

1. The Hill cipher is a multiplicative cipher.

   Multiplicative simple substitution cipher

   $$CT = key*pt \bmod 26$$

   where there are only 12 possibilities for the key: 1, 3, 5, 7, 9, 11, 15, 17, 19, 21, 23, and 25. These are the only multipliers that can be used because they are the only integers that have inverses modulo 26.

   Polygraphic substitution Hill cipher.

   $$[CT]=[key][pt] \bmod 26$$

   where the key is an $n \times n$ matrix, the plaintext and ciphertext matrices are $n \times 1$ matrices, and there must be some conditions on the key so that enciphering – multiplication by the key -- has an inverse.

2. Polygraphic ciphers, ciphers that encipher blocks of plaintext, like the Hill cipher, make frequency analysis more difficult. For simple substitution ciphers, there are only 26 frequency categories – one for each letter of the alphabet. Digraphic ciphers diffuse frequency data over $26^2 = 676$ categories, trigraphic ciphers diffuse frequency data over $26^3 = 17,576$ categories, … .

n	number of n-grams
4	456,976
5	11,881,376

Modern block ciphers typically encipher a minimum of 16 characters in a block.

3. The Hill cipher is a block cipher – every letter in the ciphertext block depends on each letter of the plaintext block. Notice in the digraphic case of the Hill cipher if one plaintext letter changes then it is likely that both ciphertext letters change. For example,

$$\begin{bmatrix}3 & 7\\ 5 & 12\end{bmatrix}\begin{bmatrix}8\\ 5\end{bmatrix}\begin{matrix}\text{h}\\ \text{e}\end{matrix} = \begin{bmatrix}3\times 8+7\times 5\\ 5\times 8+12\times 5\end{bmatrix} = \begin{bmatrix}7\\ 22\end{bmatrix}\begin{matrix}\text{G}\\ \text{V}\end{matrix}$$

but

$$\begin{bmatrix}3 & 7\\ 5 & 12\end{bmatrix}\begin{bmatrix}9\\ 5\end{bmatrix}\begin{matrix}\text{i}\\ \text{e}\end{matrix} = \begin{bmatrix}3\times 9+7\times 5\\ 5\times 9+12\times 5\end{bmatrix} = \begin{bmatrix}10\\ 1\end{bmatrix}\begin{matrix}\text{J}\\ \text{A}\end{matrix}$$

Each of the output numbers depends on both of the input numbers.

4. The Hill cipher is a substitution cipher. For the digraphic Hill cipher, the substitutions can be displayed in an 26×26 array. Such an array is often called an S-box, a substitution box. Rather than calculating each digraph substitution "on the fly" the substitutions can be calculated ahead of time and stored in a lookup table. Searching a small table is much faster than calculating each substitution as needed. On the following page is the S-box that results from using $\begin{bmatrix}3 & 7\\ 5 & 12\end{bmatrix}$ as the key. Such a box might be called an 2×2 S-box because input is a two-letter block and output is a two-letter block.

	a 1	b 2	c 3	d 4	e 5	f 6	g 7	h 8	i 9	j 10	k 11	l 12	m 13	n 14	o 15	p 16	q 17	r 18	s 19	t 20	u 21	v 22	w 23	x 24	y 25	z 26
a 1	JQ	QD	YO	EA	LM	SY	ZK	GW	NI	VV	BG	IS	PE	WQ	DC	KO	RA	YM	FY	MK	TW	AI	HV	OG	VS	CE
b 2	MV	TH	AT	HF	OR	VD	CP	JB	QN	YZ	EL	LX	SJ	ZV	GH	NT	UF	BR	ID	PP	WB	DN	KZ	RL	YX	FJ
c 3	PA	XM	DY	KK	RW	YI	FU	MG	TS	AE	HQ	OC	VO	CA	JM	QY	XK	EW	LI	SU	ZG	GS	NE	UQ	BC	IO
d 4	TF	ZR	GD	NP	UB	BN	IZ	PL	XY	DJ	KV	RH	YT	FF	MR	TD	AP	HB	ON	VZ	CL	JX	QJ	XV	EH	LT
e 5	VK	CW	JI	QU	XG	ES	LE	SQ	ZC	GO	NA	UM	BY	IL	PW	WI	DU	KG	RS	YE	FQ	MC	TO	AA	HM	OY
f 6	YP	FB	MN	TZ	AL	HX	OJ	VV	CH	JT	QF	XR	ED	LP	SB	ZN	GZ	NL	UX	BJ	IV	PH	WT	DF	KR	RD
g 7	BU	IG	PS	WE	DQ	KC	RO	YA	FM	MY	SK	AW	HI	OU	VG	CS	JE	QQ	XC	EO	LA	SM	ZY	GK	NW	US
h 8	EZ	LL	SX	ZJ	GV	NH	UT	BP	IR	PD	WP	DB	KN	RZ	YL	FX	MJ	TV	AH	HT	OF	VR	CD	JP	QB	XN
i 9	HE	OQ	VC	CO	JA	QM	XY	EK	LW	SI	ZU	GG	NS	UE	BQ	IC	PO	WA	DM	KY	RK	YW	FI	MU	TG	AS
j 10	KJ	RV	YH	FT	MF	TR	AD	HP	OB	VN	CZ	JL	QX	XJ	EV	LH	ST	ZF	GR	ND	UP	BB	IN	PZ	WL	DX
k 11	NO	UA	BM	IY	PK	WW	DI	KU	RG	YS	FE	MQ	TC	AO	HA	OM	VY	CK	JW	QI	XU	EG	LS	SE	ZQ	GC
l 12	QT	XF	ER	LD	SP	ZB	GN	NZ	UL	BX	IJ	PV	WH	DT	KF	RR	YD	FP	MB	TN	AZ	HL	OX	VJ	CV	JH
m 13	TY	AK	HW	OI	VU	CG	JS	QE	XQ	EC	LO	SA	ZM	GY	NK	UW	BI	IU	PG	WS	DE	KQ	RC	YO	FA	MM
n 14	WD	DP	KB	RN	YZ	FL	MX	TJ	AV	HH	OT	VF	CR	JD	QP	XB	EN	LZ	SL	ZX	GJ	NV	UH	BT	IF	PR
o 15	ZI	GU	NG	US	BE	IQ	PC	WO	DA	KM	RY	YK	FW	MI	TU	AG	HS	OE	VQ	CC	JO	QA	XM	EY	LK	SW
p 16	CN	JZ	QL	XX	EJ	LV	SH	ZT	GF	NR	UD	BP	IB	PN	WZ	DL	KX	RJ	YV	FH	MT	TG	AR	HD	OP	VB
q 17	FS	ME	TQ	AC	HO	OA	VM	CY	JK	QW	XI	EU	LG	SS	ZE	GQ	NC	UO	BA	IM	PY	WK	DW	KI	RU	YG
r 18	IX	PJ	WV	DH	KT	RF	YR	FD	MP	TB	AN	HZ	OL	VY	CJ	JV	QH	XT	EF	LR	SD	ZP	GB	NN	UZ	BL
s 19	LC	SO	ZA	GM	NY	UK	BW	II	PU	WG	DS	KE	RQ	YC	FO	MA	TM	AY	HK	OW	VI	CU	JG	QS	XE	EQ
t 20	OH	VT	CF	JR	QD	XP	EB	LN	SZ	ZL	GX	NJ	UV	BH	IT	PF	WR	DD	KP	RB	YN	FZ	ML	TX	AJ	HV
u 21	RC	YY	FK	MW	TI	AU	HG	OS	VE	CQ	JC	QO	XA	EM	LY	SK	ZW	GI	NU	UG	BS	IE	PQ	WC	DO	KA
v 22	UR	BD	IP	PB	WN	DZ	KL	RX	YJ	FV	MH	TT	AF	HR	OD	VP	CB	JN	QZ	XL	EX	LJ	SV	ZH	GT	NF
w 23	XW	EI	LU	SG	ZS	GE	NQ	UC	BO	IA	PM	WY	DK	KW	RI	YU	FG	MS	TE	AQ	HC	OO	VA	CM	JY	QK
x 24	AB	HN	OZ	VL	CX	JJ	QV	XH	ET	LF	SR	ZD	GP	NB	UN	BZ	IL	PX	WJ	DV	KH	RT	YF	FR	MD	TP
y 25	DG	KS	RE	YQ	FC	MO	UA	AM	HY	OK	VW	CI	JU	QG	XS	EE	LQ	SC	ZO	GA	NM	UY	BK	IW	PI	WU
z 26	GL	NX	UJ	BV	IH	PT	WF	DR	KD	RP	YB	FN	MZ	TL	AX	HJ	OV	VH	CT	JF	QR	XD	EP	LB	SN	ZZ

Decryption: Conditions on the Key

Just like for the multiplicative ciphers, we cannot use all matrices as keys because we cannot undo the multiplication for all matrices.

To go from plaintext to ciphertext we multiply by the key:

$$\begin{bmatrix}3 & 7\\ 5 & 12\end{bmatrix}\begin{bmatrix}8\\ 5\end{bmatrix}\begin{matrix}\text{h}\\ \text{e}\end{matrix} \equiv \begin{bmatrix}7\\ 22\end{bmatrix}\begin{matrix}\text{G}\\ \text{V}\end{matrix} \bmod 26$$

Now we want to undo this; we want to find a matrix so that

$$\begin{bmatrix}? & ?\\ ? & ?\end{bmatrix}\begin{bmatrix}7\\ 22\end{bmatrix} \equiv \begin{bmatrix}8\\ 5\end{bmatrix} \bmod 26$$

i.e, we want to find a matrix $\begin{bmatrix}? & ?\\ ? & ?\end{bmatrix}$ so that

$$\begin{bmatrix}? & ?\\ ? & ?\end{bmatrix}\begin{bmatrix}3 & 7\\ 5 & 12\end{bmatrix}\begin{bmatrix}8\\ 5\end{bmatrix} \equiv \begin{bmatrix}8\\ 5\end{bmatrix} \bmod 26$$

We want $\begin{bmatrix}? & ?\\ ? & ?\end{bmatrix}\begin{bmatrix}3 & 7\\ 5 & 12\end{bmatrix}$ to leave $\begin{bmatrix}8\\ 5\end{bmatrix}$ unchanged.

Matrix Inverse

The matrix we are looking for is called the inverse of $\begin{bmatrix} 3 & 7 \\ 5 & 12 \end{bmatrix}$ and is denoted $\begin{bmatrix} 3 & 7 \\ 5 & 12 \end{bmatrix}^{-1}$.

It is easy to verify that $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \begin{bmatrix} \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{bmatrix}$.

The product $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ which is called the **identity matrix** because the effect of multiplying a matrix by it is to leave the other matrix unchanged. (It is like multiplying a number by 1.)

Notice that to calculate the inverse of the matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ we must be able to divide by $ad - bc$; i.e., we must have a multiplicative inverse for $ad - bc$. Because we are working modulo 26, that means that $ad - bc$ must be one of 1, 3, 5, 7, 9, 11, 15, 17, 19, 21, 23, or 25. Otherwise, the multiplication cannot be undone; encryption cannot be undone.

Determinant

$ad - bc$ is called the **determinant** of $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Notice that the determinant of a 2×2 matrix is just the product down the upper left to lower right diagonal minus the product down the upper right to lower left diagonal. For a matrix to have an inverse modulo 26, the determinant of the matrix must be 1, 3, 5, 7, 9, 11, 15, 17, 19, 21, 23, or 25 modulo 26. To be able to undo multiplication by a matrix modulo 26, the determinant of the matrix must be 1, 3, 5, 7, 9, 11, 15, 17, 19, 21, 23, or 25 modulo 26. For a matrix to be a key for a Hill cipher, the determinant of the matrix must be 1, 3, 5, 7, 9, 11, 15, 17, 19, 21, 23, or 25 modulo 26.

The determinant of $\begin{bmatrix} 3 & 7 \\ 5 & 12 \end{bmatrix}$ is $3 \times 12 - 7 \times 5 = 1 \equiv 1 \bmod 26$. So, the inverse of $\begin{bmatrix} 3 & 7 \\ 5 & 12 \end{bmatrix}$ is

$$\begin{bmatrix} 3 & 7 \\ 5 & 12 \end{bmatrix}^{-1} = \begin{bmatrix} 12 & -7 \\ -5 & 3 \end{bmatrix} \equiv \begin{bmatrix} 12 & 19 \\ 21 & 3 \end{bmatrix} \bmod 26.$$

This is a special case because the determinant is 1.

Here is an example of finding the inverse of a 2×2 matrix when the determinant is not 1:

The determinant of $\begin{bmatrix} 9 & 4 \\ 5 & 7 \end{bmatrix}$ is $9 \times 7 - 4 \times 5 = 63 - 20 = 43 \equiv 17 \bmod 26$.

Because 17 has a multiplicative inverse modulo 26, this matrix has an inverse. The inverse of the matrix is

$$\begin{bmatrix} \frac{7}{17} & \frac{-4}{17} \\ \frac{-5}{17} & \frac{9}{17} \end{bmatrix} \bmod 26 .$$

Dividing by 17 modulo 26 is the same as multiplying by the multiplicative inverse of 17 modulo 26. Recall that the multiplicative inverse of 17 is 23 modulo 26. So, the inverse of the matrix is

$$\begin{bmatrix} \frac{7}{17} & \frac{-4}{17} \\ \frac{-5}{17} & \frac{9}{17} \end{bmatrix} \bmod 26 \equiv \begin{bmatrix} 7\times 23 & -4\times 23 \\ -5\times 23 & 9\times 23 \end{bmatrix} \bmod 26$$

$$\equiv \begin{bmatrix} 161 & -92 \\ -115 & 207 \end{bmatrix} \bmod 26 \equiv \begin{bmatrix} 5 & 12 \\ 15 & 25 \end{bmatrix} \bmod 26$$

Calculating the determinant of an $n\times n$ matrix with $n > 2$ is more difficult. The pattern used for a 2×2 matrix is a very special case. Usually calculators and computer algebra systems are able to calculate determinants.

Similarly, calculating the inverse of an $n\times n$ matrix with $n > 2$ differs from calculating the inverse of a 2×2 matrix. Again, usually calculators and computer algebra systems are able to calculate inverses.

Decryption

Encrypting the plaintext message

Herbert Yardley wrote The American Black Chamber

using the key $\begin{bmatrix} 3 & 7 \\ 5 & 12 \end{bmatrix}$ results in the ciphertext

GVPJKGAJYMRHHMMSCCYEGVPEKGVCWQLXXOBMEZAKKG

We use the inverse of the key $\begin{bmatrix} 12 & 19 \\ 21 & 3 \end{bmatrix}$ to decrypt GV, which is the first digraph of the ciphertext.

$$\begin{bmatrix} 12 & 19 \\ 21 & 3 \end{bmatrix} \begin{bmatrix} 7 \\ 22 \end{bmatrix} \begin{matrix} \text{G} \\ \text{V} \end{matrix} \equiv \begin{bmatrix} 8 \\ 5 \end{bmatrix} \begin{matrix} \text{h} \\ \text{e} \end{matrix} \bmod 26$$

In a similar manner, we can decrypt the remainder of the ciphertext.

Key Construction

So, it is necessary that the determinant of the key be one of 1, 3, 5, 7, 9, 11, 15, 17, 19, 21, or 23 modulo 26.

How likely is that to happen for a 2×2 matrix?

$$\begin{bmatrix} \square & \square \\ \square & \square \end{bmatrix}$$

There are $26^4 = 456{,}976$ 2×2 matrices with entries 1, 2, 3, … , 26. The number that have inverses is 157,248. So, the probability of "drawing one at random" is

$$\frac{157248}{456976} \approx 0.34$$

about 1 in 3 have inverses.

Therefore, a reasonable way to generate a key is to generate a random 2×2 matrix and test whether it has an inverse and repeat this process until a valid key is found.

Random matrices can be generated by calculator

randmat(2, 2)

and then check the determinant

mod(det(ans(1)), 26)

or using *Mathematica*

RandomInteger[{1, 26}, {2, 2}]

and then check the determinant

Mod[Det[%], 26].

Number of trials	Probability that all are NOT invertible
1	0.67
2	0.44
3	0.296
4	0.198
5	0.132
6	0.088
7	0.059
8	0.039
9	0.026
10	0.017
11	0.011
12	0.008

It is likely in just a few trials that a key with an inverse can be found.

Known plaintext attack

It can be difficult to cryptanalyze a Hill cipher using a ciphertext only attack, but it is easy to break using a **known plaintext attack**. A known plaintext attack means that we know a bit of ciphertext and the corresponding plaintext – a crib. This is not an unusual situation. Often messages have stereotypical beginnings (e.g., *to ...*, *dear ...*) or stereotypical endings (e.g, *stop*) or sometimes it is possible (knowing the sender and receiver or knowing what is likely to be the content of the message) to guess a portion of a message.

For a 2×2 Hill cipher, if we know two ciphertext digraphs and the corresponding plaintext digraphs, we can easily determine the key or the key inverse. Assume that we know that the plaintext of our ciphertext message that begins WBVE is inma. Because WB corresponds to in $\begin{bmatrix} e & f \\ g & h \end{bmatrix}\begin{bmatrix} 23 \\ 2 \end{bmatrix}=\begin{bmatrix} 9 \\ 14 \end{bmatrix}$, and because VE

corresponds to ma $\begin{bmatrix} e & f \\ g & h \end{bmatrix}\begin{bmatrix} 22 \\ 5 \end{bmatrix}=\begin{bmatrix} 13 \\ 1 \end{bmatrix}$. This results in two sets of linear congruences modulo 26:

$$\begin{aligned} 23e &+ 2f = 9 \\ 22e &+ 5f = 13 \end{aligned}$$

and

$$\begin{aligned} 23g &+ 2h = 14 \\ 22g &+ 5h = 1 \end{aligned}$$

We solve the systems modulo 26 using *Mathematica.*

Solve[23e + 2f == 9 && 22e + 5f == 13, {e, f}, Modulus -> 26]

{{e->1,f->19}}

Solve[23g + 2h == 14 && 22g + 5h ==1, {g, h}, Modulus -> 26]

{{g->20,h->11}}

Again (with a lot less assuming) we find that the key inverse is $\begin{bmatrix} 1 & 19 \\ 20 & 11 \end{bmatrix}$.

Two More Examples of a Known Plaintext Attack

Here are two examples of cryptanalyzing a Hill cipher with a known plaintext attack. Each example is done by hand – without using *Mathematica*. In example one, there is no need to reduce the modulus; in example two the modulus must be reduced.

Example one

Ciphertext: FAGQQ ILABQ VLJCY QULAU STYTO JSDJJ PODFS ZNLUH KMOW

We are assuming that this message was encrypted using a 2×2 Hill cipher and that we have a crib. We believe that the message begins "a crib."

ac	ri
[1, 3]	[18, 9]
[6, 1]	[7, 17]
FA	GQ

We could either solve for the key or the key inverse. To solve for the key, we would solve

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} 1 \\ 3 \end{bmatrix}=\begin{bmatrix} 6 \\ 1 \end{bmatrix}$$

and

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} 18 \\ 9 \end{bmatrix}=\begin{bmatrix} 7 \\ 17 \end{bmatrix}$$

To solve for the key inverse, we would solve

$$\begin{bmatrix} e & f \\ g & h \end{bmatrix}\begin{bmatrix} 6 \\ 1 \end{bmatrix}=\begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

and

$$\begin{bmatrix} e & f \\ g & h \end{bmatrix}\begin{bmatrix} 7 \\ 17 \end{bmatrix}=\begin{bmatrix} 18 \\ 9 \end{bmatrix}$$

We will solve for the key.

$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} 1 \\ 3 \end{bmatrix}=\begin{bmatrix} 6 \\ 1 \end{bmatrix}$ represents two linear equations modulo 26:

$$\begin{aligned} a \quad +3b &= 6 \\ c \quad +3d &= 1 \end{aligned}$$

and $\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} 18 \\ 9 \end{bmatrix}=\begin{bmatrix} 7 \\ 17 \end{bmatrix}$ represents

$$\begin{aligned} 18a \quad +9b &= 7 \\ 18c \quad +9d &= 17 \end{aligned}$$

Now we solve the following linear congruences mod 26.

$$\begin{cases} a \quad +3b = 6 \\ 18a \quad +9b = 7 \end{cases} \text{ and } \begin{cases} c \quad +3d = 1 \\ 18c \quad +9d = 17 \end{cases}$$

We will solve the pair of congruences $\begin{cases} a \quad +3b = 6 \\ 18a \quad +9b = 7 \end{cases}$ first.

To eliminate an unknown, multiply congruence 1 by 3

$$\begin{cases} 3a & +9b & = & 18 \\ 18a & +9b & = & 7 \end{cases}$$

and subtract congruence 2 from congruence 1.

$$-15a \quad = \quad 11$$

Modulo 26, -15 is 11.

$$11a \quad = \quad 11$$

Divide by 11 to obtain a.

$$a = 1$$

Now substitute this in congruence 1.

$$1 + 3b = 6$$

$$3b = 5$$

The multiplicative inverse of 3 is 9 modulo 26.

$$b = 9 \times 3b = 9 \times 5 = 45 = 19 \bmod 26$$

So, the key looks like

$$\begin{bmatrix} 1 & 19 \\ c & d \end{bmatrix}$$

Now solve the system $\begin{cases} c & +3d & = & 1 \\ 18c & +9d & = & 17 \end{cases}$

$$\begin{cases} 3c & +9d & = & 3 \\ 18c & +9d & = & 17 \end{cases}$$

$$15c = 14$$

$$c = 7 \times 15c = 7 \times 14 = 98 = 20 \bmod 26$$

$$20 + 3d = 1$$

$$3d = -19 = 7 \bmod 26$$

$$d = 9 \times 3d = 9 \times 7 = 63 = 11 \bmod 26$$

The key is $\begin{bmatrix} 1 & 19 \\ 20 & 11 \end{bmatrix}$.

Example two

We are assuming that we have a ciphertext message was that encrypted using a 2×2 Hill cipher and that we have a crib.

We believe that ciphertext UKJN corresponds to plaintext word.

wo	rd
[23, 15]	[18, 4]
[21, 11]	[10, 14]
UK	JN

The two systems of congruences are:

$$\begin{cases} 23a & +15b & = & 21 \\ 18a & +4b & = & 10 \end{cases} \text{ and } \begin{cases} 23c & +15d & = & 11 \\ 18c & +4d & = & 14 \end{cases}$$

We will solve the system on the left.

To eliminate an unknown, multiply congruence number 1 by 4 and congruence number 2 by 15 both modulo 26.

$$\begin{cases} 14a & +8b & = & 6 \\ 10a & +8b & = & 20 \end{cases}$$

Subtract the second congruence from the first.

$$4a = -14 = 12 \bmod 26$$

This congruence corresponds to the equation $4a = 12 + 26k$, $4a$ is 12 plus a multiple of 26. Notice that 2 divides the coefficient of a, the constant 12, and the modulus 26. We reduce the modulus by dividing by 2.

$$2a = 6 + 13k$$

and we have a congruence modulo 13.

$$2a = 6 \bmod 13$$

This congruence does not have a common factor among the coefficient, the constant, and the modulus.

Here are the multiplicative inverses of the integers modulo 13:

Number	1	2	3	4	5	6	7	8	9	10	11	12
Multiplicative inverse	1	7	9	10	8	11	2	5	3	4	6	12

To find a, multiply $2a = 6 \bmod 13$ by the multiplicative inverse of 2, which is 7.

$$a = 7 \times 2a = 7 \times 6 = 42 = 3 \bmod 13$$

So, a is 3 modulo 13. But, there are two integers mod 26 that are 3 mod 13, namely, 3 and 3 + 13 = 16. So, there are two possible values for a.

If $a = 3$,

$$18 \times 3 = 4b = 10$$

$$54 + 4b = 10$$

$$2 + 4b = 10$$

$$4b = 8 \bmod 26$$

$$2b = 4 \bmod 13$$

$$b = 7 \times 2b = 7 \times 4 = 26 = 2 \bmod 13$$

So, b=2 or b = 2+13 = 15 modulo 16.

If $a = 16$,

$$18 \times 16 + 4b = 10$$

$$288 + 4b = 10$$

$$2 + 4b = 10$$

which yields the same solutions for b.

Here are the 4 possible solutions for a and b.

$$\begin{array}{ll} a = 3 & b = 2 \\ a = 3 & b = 15 \\ a = 16 & b = 2 \\ a = 16 & b = 15 \end{array}$$

Now solve $\begin{cases} 23c + 15d = 11 \\ 18c + 4d = 14 \end{cases}$.

$$\begin{cases} 14a + 8b = 18 \\ 10a + 8b = 2 \end{cases}$$

$$4c = 16 \bmod 26$$

$$2c = 8 \bmod 13$$

$$c = 7 \times 2c = 14c = 7 \times 8 = 56 = 4 \bmod 13$$

So, $c = 4$ or $c = 4 + 13 = 17$ modulo 26.

If $c = 4$,

$$18 \times 4 + 4d = 14$$

$$20 + 4d = 14$$

$$4d = -6 = 20 \bmod 26$$

$$2d = 10 \bmod 13$$

$$d = 7 \times 2d = 7 \times 10 = 5 \bmod 13$$

So, $d = 5$ or $d = 5 + 13 = 18$ modulo 26.

If $c = 17$,

$$18 \times 17 + 4d = 14$$

$$20 + 4d = 14$$

and we are led to the same solutions for d.

$$\begin{array}{ll} c = 4 & d = 5 \\ c = 4 & d = 18 \\ c = 17 & d = 5 \\ c = 17 & d = 18 \end{array}$$

There are 16 possible 2×2 matrices that could be the key.

$$\begin{bmatrix} 3 & 2 \\ 4 & 5 \end{bmatrix} \quad \begin{bmatrix} 3 & 2 \\ 4 & 18 \end{bmatrix} \quad \begin{bmatrix} 3 & 2 \\ 17 & 5 \end{bmatrix} \quad \begin{bmatrix} 3 & 2 \\ 17 & 18 \end{bmatrix}$$
$$\begin{bmatrix} 3 & 15 \\ 4 & 5 \end{bmatrix} \quad \begin{bmatrix} 3 & 15 \\ 4 & 18 \end{bmatrix} \quad \begin{bmatrix} 3 & 15 \\ 17 & 5 \end{bmatrix} \quad \begin{bmatrix} 3 & 15 \\ 17 & 18 \end{bmatrix}$$
$$\begin{bmatrix} 16 & 2 \\ 4 & 5 \end{bmatrix} \quad \begin{bmatrix} 16 & 2 \\ 4 & 18 \end{bmatrix} \quad \begin{bmatrix} 16 & 2 \\ 17 & 5 \end{bmatrix} \quad \begin{bmatrix} 16 & 2 \\ 17 & 18 \end{bmatrix}$$
$$\begin{bmatrix} 16 & 15 \\ 4 & 5 \end{bmatrix} \quad \begin{bmatrix} 16 & 15 \\ 4 & 18 \end{bmatrix} \quad \begin{bmatrix} 16 & 15 \\ 17 & 5 \end{bmatrix} \quad \begin{bmatrix} 16 & 15 \\ 17 & 18 \end{bmatrix}$$

First, calculate the determinant of each. Any matrix that does not have an invertible determinant modulo 26 (i.e., the determinant is not one of 1, 3, 5, 7, 9, 11, 15, 17, 19, 21, 23, 25 modulo 26) can be eliminated. Then try to decipher the messages with each of the remaining messages. The matrix that yields plaintext is the key.

To break a Hill cipher with a 2×2 key requires determining four entries – the four entries of the key or the four entries of the key inverse. We can do that if we know the correspondence between plaintext and ciphertext for two (independent) digraphs because the correspondences will permit us to set up two systems of congruences – each system has two congruences of two unknowns.

To break a Hill cipher with a $n \times n$ key requires determining n^2 entries – the n^2 entries of the key or the n^2 entries of the key inverse. We can do that if we know the correspondence between plaintext and ciphertext for n n-graphs because the correspondences will permit us to set up n systems of congruences – each system has n congruences of n unknowns.

The reason that we can solve these systems of congruences is because they are linear. The solutions of linear systems of equations of congruences is well-understood. The technique used to solve systems of linear equations Gaussian elimination is a very efficient algorithm.

Re-encryption

Encrypting with a Hill cipher and re-encrypting with another key of the same size does not improve security.

For example, if we encrypted digraphs with a Hill cipher using the key $\begin{bmatrix} 3 & 7 \\ 5 & 12 \end{bmatrix}$ (which has determinant 1 modulo 26) and then encrypted that ciphertext using a Hill cipher with key $\begin{bmatrix} 9 & 4 \\ 5 & 7 \end{bmatrix}$ (which has determinant 17 modulo 26), the result would be the same as encrypting once with the key

$$\begin{bmatrix} 9 & 4 \\ 5 & 7 \end{bmatrix}\begin{bmatrix} 3 & 7 \\ 5 & 12 \end{bmatrix} = \begin{bmatrix} 47 & 111 \\ 50 & 119 \end{bmatrix} = \begin{bmatrix} 21 & 7 \\ 24 & 15 \end{bmatrix} \text{ modulo } 26.$$

Because "the determinant of a product is the product of the determinants" (even modulo 26), the determinant of $\begin{bmatrix} 21 & 7 \\ 24 & 15 \end{bmatrix}$ is $17 \times 1 = 17 \text{ modulo } 26$; so, it is a valid Hill cipher key.

The point is, you have only one shot at using a Hill cipher – re-encrypting does not improve security.

Other Hill Ciphers

Hill's papers contain techniques that are much more secure than the technique that we have called the Hill cipher. Hill's papers include ciphers that are nonlinear.

One technique used by Hill is to do a (nonlinear) simple substitution cipher – a permutation -- prior to the matrix multiplication. Hill uses the following substitutions:

a b c d e f g h i j k l m n o p q r s t u v w x y z

5 23 2 20 10 15 8 4 18 25 0 16 13 7 3 1 19 6 12 24 21 17 14 22 11 9

For example, th becomes 24 4 and then

$$\begin{bmatrix} 3 & 7 \\ 5 & 12 \end{bmatrix} \begin{bmatrix} 24 \\ 4 \end{bmatrix} = \begin{bmatrix} 22 \\ 12 \end{bmatrix}$$

V (22) L (12).

Another technique used by Hill is similar to what we did when we went from the multiplicative cipher ($C = mp$) to the affine cipher ($C = mp + b$) by adding a shift. For example,

$$\begin{bmatrix} 3 & 7 \\ 5 & 12 \end{bmatrix}\begin{bmatrix} 8 \\ 5 \end{bmatrix}\begin{matrix} \text{h} \\ \text{e} \end{matrix} + \begin{bmatrix} 6 \\ 20 \end{bmatrix} = \begin{bmatrix} 13 \\ 16 \end{bmatrix}\begin{matrix} \text{M} \\ \text{P} \end{matrix}.$$

Hill's 1929 and 1931 papers include other generalizations of the Hill cipher.

Involutory Keys

Hill ciphers have two keys: one key is used for encryption and a second key (the key inverse) is used for decryption. (This is true for all of the ciphers that we have studied; however, the relationship between the key and key inverse seems less obvious for the Hill cipher.) Of course, anyone who knows some elementary linear algebra can construct the key inverse from the key, but the encryption and decryption keys are not the same – except in certain cases. Hill, in his second paper, discusses using involutory matrices (matrices that are self-inverse) as keys.

$$\begin{bmatrix} 0 & 1 & 25 \\ 4 & 22 & 4 \\ 3 & 22 & 4 \end{bmatrix} \text{ is involutory.}$$

Using involutory keys would make encryption and decryption completely symmetric, but this significantly restricts the number of keys.

The Hill Cipher

Hill's cipher is overtly an algebraic cipher – it might have been the first. And, Hill's cipher generalizes to any dimension and, therefore, moved cryptography beyond encrypting digraphs to encrypting blocks.

It appears that the weakness of Hill's cipher – that it is linear and, therefore, subject to a known plaintext attack -- was known early. So, the only known use of Hill's cipher was to encrypt Navy radio station call signs.

Ciphertext Attack

Here is a ciphertext that is known to be enciphered with a Hill cipher.

```
wbvec itxwb mphsr hytyw gmqdq egxyf yncta zdkyi eenin zkygh
yntgb pbpkl azfgy ikkru drzcp aaaci fueqg ywbuu urozm vfgmy
vkwwo zbpyn ezsbg jfynz yvmeo zctiu ghfgu aekds ayicc tkrus
xgbpz cufve lvsjg lklls vefyt onmdk
```

The first thing to be determined would be the size of the blocks. If the key were an $n \times n$ matrix, then n must divide the number of letters in the ciphertext. This ciphertext has 180 letters. There are many possibilities for n, but let us assume that it was encrypted using a 2×2 key. (That's a really good assumption.)

Because such a key encrypts digraphs, we might begin by looking at digraph frequencies.

Here are the digraphs that appear more than once and their frequencies:

Digraph	Frequency
bp	4
yn	4
ct	3
fg	3
fy	3
oz	3
ve	3
wb	3
yi	3
zc	3
aa	2
az	2
ci	2
dk	2
gb	2
gh	2
gm	2
gy	2
hy	2
kl	2
kr	2
ky	2
nz	2
ru	2
uu	2
yt	2
yv	2
yw	2

If we're lucky the most common plaintext digraph th will correspond to (one of) the most common ciphertext digraph(s). BP and YN each appear 4 times in the ciphertext. Let's assume that ciphertext BP corresponds to plaintext th. (Another really good assumption.)

We could try to determine the key or the key inverse. Because we are trying to determine the plaintext, let's try to directly determine the key inverse. We want to find a 2×2 matrix $\begin{bmatrix} e & f \\ g & h \end{bmatrix}$ that is the inverse of the key. If we are correct that B(2)P(16) corresponds to t(20)h(8), then

$$\begin{bmatrix} e & f \\ g & h \end{bmatrix}\begin{bmatrix} 2 \\ 16 \end{bmatrix}=\begin{bmatrix} 20 \\ 8 \end{bmatrix}$$

This corresponds to two linear equations:

$$\begin{aligned} 2e &+ 16f &&= 20 \\ 2g &+ 16h &&= 8 \end{aligned}$$

Because this Hill cipher (we assume) encrypts digraphs, the key inverse is a 2×2 matrix. The key inverse has $2^2 = 4$ entries e, f, g, and h that must be determined. We would like to have four equations – two involving e and f and two involving g and h.

If we knew another plaintext/ciphertext digraph correspondence, we would have the other two equations that we need. Perhaps, the next most common ciphertext digraph YN corresponds to the next most common plaintext digraph he. (But, it doesn't.)

We could try assuming that YN corresponds to another common digraph, but here is another technique.

The most common letter that follows plaintext th is e. We might examine the digraphs that follow BP in the ciphertext and assume that the next ciphertext digraph corresponds to plaintext e_. We notice that we have BP KL, BP BP, BP YN, and BP ZC. If we are correct that BP corresponds to th, the second pair of digraphs corresponds to plaintext th th. In each of the other cases, we will assume that the two ciphertext digraphs correspond to th e_. Making this assumption, we should be correct more than half the time.

So, if KL corresponds to e_, $\begin{bmatrix} e & f \\ g & h \end{bmatrix}\begin{bmatrix} 11 \\ 12 \end{bmatrix} = \begin{bmatrix} 5 \\ * \end{bmatrix}$ which yields the equation $11e+12f=5$. If YN corresponds to e_, $\begin{bmatrix} e & f \\ g & h \end{bmatrix}\begin{bmatrix} 25 \\ 14 \end{bmatrix} = \begin{bmatrix} 5 \\ * \end{bmatrix}$ which yields the equation $25e+14f=5$. If ZC corresponds to e_, $\begin{bmatrix} e & f \\ g & h \end{bmatrix}\begin{bmatrix} 26 \\ 3 \end{bmatrix} = \begin{bmatrix} 5 \\ * \end{bmatrix}$ which yields the equation $26e+3f=5$.

Each of these can be solved simultaneously with, $12e+16f=20$ which was obtained by assuming that BP corresponds to th. All of the solving, however, is to be done modulo 26. We may use whatever techniques we know for solving systems of linear equations provided that we divide only when division is possible – we can divide by only 1, 3, 5, 7, 9, 11, 15, 17, 19, 21, 23, and 25. We will use *Mathematica* to solve the equations.

Solve[2e+16f==20&&11e+12f==5,{e,f},Modulus ->26]

{{e->1,f->6+13 C[1]}}

Solve[2e+16f==20&&25e+14f==5,{e,f},Modulus -> 26]

{{e->1,f->6+13 C[1]}}

Solve[2e+16f==20&&26e+3f==5,{e,f},Modulus -> 26]

{{e->1+13 C[1],f->19}}

Each system of congruences has two solutions modulo 26. $e = 1$ and $f = 19$ is common to all of the pairs of solutions. That would happen if in each of these three cases th were followed by e_. Let us assume that is the case. (That's another really good assumption.) We could later try the other possibilities if needed.

So, we believe that the key inverse is $\begin{bmatrix} 1 & 19 \\ g & h \end{bmatrix}$.

We have one more congruence: $2g+16h=8 \bmod 26$. It is possible to solve a congruence of the form $ax+by=c \bmod n$ provided that the greatest common divisor of a, b, and n also divides c. In our case, the greatest common divisor of a=2, b=16, and n=26 is 2 which does divide c=8. It is necessary to reduce the modulus; remove the factor of 2 to get $g+8h=4 \bmod 13$. Then rearrange the terms to get $g=4-8h \bmod 13$. Modulo 13, the possible values of h are 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, and 12.

For example, if $h = 3$, $g = 6$ mod 13. But, we are ultimately interested in what happens modulo 26. 6 and 6 + 13 = 19 are congruent mod 13, but they are not congruent mod 26. So, each solution mod 13 becomes two solutions mod 26.

h	g mod 13	g mod 26
0	4	4, 17
1	9	9, 22
2	1	1, 10
3	6	6, 19
4	11	11, 24
5	3	3, 22
6	8	8, 21
7	0	0, 13
8	5	5, 18
9	10	10, 23
10	2	2, 15
11	7	7, 20
12	12	12, 25

The determinant of the key inverse must be one of 1, 3, 5, 7, 9, 11, 15, 17, 19, 21, 23, or 25 mod 26. So, try each of these pairs of g and h in $\begin{bmatrix} 1 & 19 \\ g & h \end{bmatrix}$ and calculate the determinant mod 26. Again, we use *Mathematica*.

In[8]:= Mod[Det[{{1, 19}, {4, 0}}], 26]

Out[8]= 2

Mod[Det[{{1, 19}, {17, 0}}], 26]

15

Mod[Det[{{1, 19}, {9, 1}}], 26]

12

Mod[Det[{{1, 19}, {22, 1}}], 26]

25

Mod[Det[{{1, 19}, {1, 2}}], 26]

9

Mod[Det[{{1, 19}, {10, 2}}], 26]

20

Mod[Det[{{1, 19}, {6, 3}}], 26]

19

Mod[Det[{{1, 19}, {19, 3}}], 26]

6

Mod[Det[{{1, 19}, {11, 4}}], 26]

3

Mod[Det[{{1, 19}, {24, 4}}], 26]

16

Mod[Det[{{1, 19}, {3, 5}}], 26]

0

Mod[Det[{{1, 19}, {22, 5}}], 26]

3

Mod[Det[{{1, 19}, {8, 6}}], 26]

10

Mod[Det[{{1, 19}, {21, 6}}], 26]

23

Mod[Det[{{1, 19}, {0, 7}}], 26]

7

Mod[Det[{{1, 19}, {13, 7}}], 26]

20

Mod[Det[{{1, 19}, {5, 8}}], 26]

17

Mod[Det[{{1, 19}, {18, 8}}], 26]

4

Mod[Det[{{1, 19}, {10, 9}}], 26]

1

Mod[Det[{{1, 19}, {23, 9}}], 26]

14

Mod[Det[{{1, 19}, {2, 10}}], 26]

24

Mod[Det[{{1, 19}, {15, 10}}], 26]

11

Mod[Det[{{1, 19}, {7, 11}}], 26]

8

Mod[Det[{{1, 19}, {20, 11}}], 26]

21

Mod[Det[{{1, 19}, {12, 12}}], 26]

18

Mod[Det[{{1, 19}, {25, 12}}], 26]

5

The possible key inverses are $\begin{bmatrix} 1 & 19 \\ 17 & 0 \end{bmatrix}$, $\begin{bmatrix} 1 & 19 \\ 22 & 1 \end{bmatrix}$, $\begin{bmatrix} 1 & 19 \\ 1 & 2 \end{bmatrix}$, $\begin{bmatrix} 1 & 19 \\ 6 & 3 \end{bmatrix}$, $\begin{bmatrix} 1 & 19 \\ 11 & 54 \end{bmatrix}$, $\begin{bmatrix} 1 & 19 \\ 22 & 5 \end{bmatrix}$, $\begin{bmatrix} 1 & 19 \\ 21 & 6 \end{bmatrix}$, $\begin{bmatrix} 1 & 19 \\ 0 & 7 \end{bmatrix}$, $\begin{bmatrix} 1 & 19 \\ 5 & 8 \end{bmatrix}$, $\begin{bmatrix} 1 & 19 \\ 10 & 9 \end{bmatrix}$, $\begin{bmatrix} 1 & 19 \\ 15 & 10 \end{bmatrix}$, $\begin{bmatrix} 1 & 19 \\ 20 & 11 \end{bmatrix}$, and $\begin{bmatrix} 1 & 19 \\ 25 & 12 \end{bmatrix}$.

We have reduced the problem to checking 13 possible key inverses. We try to decrypt the ciphertext with each possible inverse. $\begin{bmatrix} 1 & 19 \\ 20 & 11 \end{bmatrix}$ is the correct key inverse.