9. Distance measures

9.1 Classical information measures

How similar/close are two probability distributions \( \{p_x\}, \{q_x\} \)?

**Trace distance**

\[
D(p_x, q_x) = \frac{1}{2} \sum_x |p_x - q_x|
\]

**Fidelity**

\[
F(p_x, q_x) = \sum_x \sqrt{p_x q_x}
\]

**Example:** Flipping two coins, one fair one biased

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Head</td>
<td>Tail</td>
</tr>
<tr>
<td>(p_1 = 1/2)</td>
<td>(p_2 = 1/2)</td>
</tr>
<tr>
<td>(q_1 = 1/2 + \epsilon)</td>
<td>(q_2 = 1/2 - \epsilon)</td>
</tr>
</tbody>
</table>
Trace distance

\[ D(p_x, q_x) = \frac{1}{2} (|p_1 - q_1| + |p_2 - q_2|) = \epsilon \]

Fidelity

\[ F(p_x, q_x) = \sqrt{p_1 q_1} + \sqrt{p_2 q_2} = \frac{1}{2} \left( \sqrt{1 - 2\epsilon} + \sqrt{1 + 2\epsilon} \right) \]

Two outcomes, general

\[ p_1 = p, \quad p_2 = 1 - p \]
\[ q_1 = q, \quad q_2 = 1 - q \]

This gives

\[ D(p_x, q_x) = |p - q| \]
\[ F(p_x, q_x) = \sqrt{pq} + \sqrt{(1 - p)(1 - q)} \]

\[ \implies \quad \text{for} \quad p - q \rightarrow 0 \quad D(p_x, q_x) \rightarrow 0 \]
\[ \quad \quad \quad \text{for} \quad p - q \rightarrow 0 \quad F(p_x, q_x) \rightarrow 1 \]
9.1 Quantum information measures

How similar/close are two quantum states \( \rho, \sigma \) ?

**Trace distance**

\[
D(\rho, \sigma) = \frac{1}{2} \text{tr}|\rho - \sigma|
\]

where \( |A| = \sqrt{A^\dagger A} \)

**Fidelity**

\[
F(\rho, \sigma) = \text{tr}\sqrt{\rho^{1/2}\sigma\rho^{1/2}}
\]

**Properties**

For commuting (simultaneously diagonalizable) \( \rho, \sigma \) as

\[
\rho = \sum_i r_i |i\rangle \langle i| \quad \sigma = \sum_i s_i |i\rangle \langle i|
\]

we recover the classical measures

\[
D(\rho, \sigma) = D(r_i, s_i) \quad F(\rho, \sigma) = F(r_i, s_i)
\]

**Exercise 9.6 (extended)**
Quantum operations

Quantum operations are contractive on the space of density operators

\[ D(\rho, \sigma) \geq D(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \]
\[ F(\rho, \sigma) \leq F(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \]

Relations

\[ 1 - F(\rho, \sigma) \leq D(\rho, \sigma) \leq \sqrt{1 - F(\rho, \sigma)^2} \]

Special properties

Qubit: \[ \rho = \frac{1}{2} [I + \bar{r} \cdot \bar{\sigma}], \quad \sigma = \frac{1}{2} [I + \bar{s} \cdot \bar{\sigma}] \]

\[ D(\rho, \sigma) = \frac{1}{2} |\bar{r} - \bar{s}| \]

Pure states: \[ F(|\psi\rangle, |\varphi\rangle) = |\langle \psi | \varphi \rangle| \]
11. Entropy and information

Classical information

$X$ is a random variable with probability distribution $p_1, \ldots, p_n$

$$H(X) \equiv - \sum_{x=1}^{n} p_x \log(p_x) \quad \text{Shannons entropy}$$

Shannons entropy is "amount of uncertainty" before knowing $X$, or "amount of information" gained by learning about $X$.

For two outcomes $p_1 = p, \ p_2 = 1 - p$

$$H(X) \equiv -p \log(p) - (1-p) \log(1-p) \quad \text{Binary entropy}$$

Limits

$p = 1 \quad \text{or} \quad p = 0 \quad \Rightarrow \quad H(X) = 0$

$p = 1/2 \quad \Rightarrow \quad H(X) = 1$
Two distributions

$X, Y$ are random variables with joint probability distribution $p(x, y)$

$$H(X, Y) \equiv -\sum_{x,y} p(x, y) \log[p(x, y)] \quad \text{Joint entropy}$$

Independent variables $X, Y$

$$p(x, y) = p_x p_y \implies H(X, Y) = H(X) + H(Y)$$

**Example:** Flipping a coin. $X$ - US or EU, $Y$ - head or tail

<table>
<thead>
<tr>
<th>Type</th>
<th>US</th>
<th>EU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prob.</td>
<td>$q$</td>
<td>$1-q$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Head</th>
<th>$p_A$</th>
<th>$p_E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tail</td>
<td>$1-p_A$</td>
<td>$1-p_E$</td>
</tr>
</tbody>
</table>

Variables $X, Y$ not independent
- Entropy of $X$ conditional on knowing $Y$

$$H(X|Y) \equiv H(X, Y) - H(Y)$$

*Conditional entropy*

Note that $p_y = \sum_x p(x, y)$.

For independent variables $H(X|Y) = H(X)$

- Mutual information in $X$ and $Y$

$$H(X : Y) \equiv H(X) + H(Y) - H(X, Y)$$

*Mutual information*

For independent variables $H(X : Y) = 0$
Quantum information

Density matrix, spectral decomposition

\[ \rho = \sum_x \lambda_x |x\rangle \langle x| \]

Information content in $\rho$

\[ S(\rho) \equiv -\text{tr} [\rho \log(\rho)] = -\sum_x \lambda_x \log(\lambda_x) \]

Von Neumann entropy

Note that $S(\rho) = H(X)$, for $p_x = \lambda_x$.

Maximum entropy

For a $d$-dimensional Hilbert space

\[ \rho = \frac{1}{d} \quad \Rightarrow \quad S(\rho) = \log(d) \]

Qubit

\[ \rho = \lambda |\lambda_1\rangle \langle \lambda_1| + (1 - \lambda) |\lambda_2\rangle \langle \lambda_2| \]

\[ S(\rho) \equiv -\lambda \log \lambda - (1 - \lambda) \log(1 - \lambda) \]

Binary entropy

Exercise 11.11 (parts)
Exercise 11.12
$S(\rho) \leq H(p, 1 - p)$
Composite system

Joint density matrix $\rho_{AB}$

- $S(A, B) \equiv S(\rho_{AB}) = -\text{tr}[\rho_{AB} \log(\rho_{AB})]$ \textbf{Joint entropy}

- Tensor product $\rho_{AB} = \rho_A \otimes \rho_B \implies S(A, B) = S(A) + S(B)$

- $S(A|B) \equiv S(A, B) - S(B)$ \textbf{Conditional entropy}

- $S(A : B) \equiv S(A) + S(B) - S(A, B)$ \textbf{Mutual information}
12. Quantum information theory

12.1 Quantum states and accessible information

Distinguishable states
Alice has a classical binary information source $X$, producing bits

$$x = 0 \quad \text{with probability} \quad p$$

$$x = 1 \quad \text{with probability} \quad 1 - p$$

Shannon entropy $H(X) = -p \log(p) - (1 - p) \log(1 - p)$

For every bit Alice gets she sends a qubit to Bob

for $x = 0$, she sends $|0\rangle$

for $x = 1$, she sends $|1\rangle$

Bob receives the quantum state

$$\rho = p |0\rangle \langle 0| + (1 - p) |1\rangle \langle 1|$$

If Bob makes a projection measurement in the basis $|0\rangle$, $|1\rangle$ he can reconstruct the information Alice sent

$$S(\rho) = -p \log(p) - (1 - p) \log(1 - p) = H(X)$$
Indistinguishable states

Now Alice runs out of $|1\rangle$. For every classical bit Alice gets she sends:

for $x = 0$, she sends $|0\rangle$

for $x = 1$, she sends $\cos \theta |0\rangle + \sin \theta |1\rangle$

Bob receives the quantum state

$$\rho = p |0\rangle \langle 0 | + (1 - p) \left[ \cos \theta |0\rangle + \sin \theta |1\rangle \right] \left[ \langle 0 | \cos \theta + \langle 1 | \sin \theta \right]$$

Can Bob reconstruct the information? Not perfectly, since $|0\rangle$ and $\cos \theta |0\rangle + \sin \theta |1\rangle$ are not distinguishable (orthogonal).

**Problem**: How much information about $X$ can Bob retrieve by performing a measurement yielding a random variable $Y$?

Quantified by mutual information

$$H(X : Y) = H(X) + H(Y) - H(X, Y)$$

We know $H(X) = -p \log(p) - (1 - p) \log(1 - p)$. The measurement determines $H(Y), H(X, Y)$. 
Projection measurement
Bob projects in the basis $|0\rangle, |1\rangle$

**Derivation** $H_{proj}(X : Y)$

Generalized measurement
Bob performs a POVM measurement

**Derivation** $H_{POVM}(X : Y)$

We see that the mutual information $H_{proj}(X : Y) \geq H_{POVM}(X : Y)$
Accessible information

**Definition:**

*Accessible information*  $H_{acc}(A : B)$: largest mutual information for all possible measurement schemes.

**Holevo bound**

Suppose Alice prepares a state $\rho_x$ where $x = 1, \ldots, n$ with probabilities $p_1, \ldots, p_n$ and sends the state to Bob. The accessible information for Bob is

$$H_{acc}(A : B) = S(\rho) - \sum_x p_x S(\rho_x) \equiv \chi$$

where $\rho = \sum_x p_x \rho_x$ (here no proof).

For our case this gives the accessible information

**Derivation**  $\chi$

Well above projection and POVM schemes!
12.2 Data compression

Classical information

**Question:** How much can classical information be compressed, transmitted and then decompressed without losses?

As a model for an information source we take a sequence of random variables $X_1, X_2, \ldots, X_n$ independently and identically distributed, $p_1, \ldots, p_d$.

Such an *i.i.d.* model is generally not true but typically works for large $n$.

**Example:** We consider a binary source which emits bits

\begin{align*}
x &= 0 \quad \text{with probability } \ p \\
x &= 1 \quad \text{with probability } \ 1 - p
\end{align*}
Key idea: typical sequence

The sequence of $n$ bits are parted into typical and atypical, e.g. for $p = 1/2$

\[
\begin{align*}
00101001101011010011001001110001 & \quad \text{typical} \\
00000100000010000000010000000000 & \quad \text{atypical}
\end{align*}
\]

Typical sequences have $\approx np$ bits 0 and $\approx n(1 - p)$ bits 1.

The probability for a given typical sequence is

\[
p(x_1, \ldots, x_n) = p(x_1)p(x_2)\ldots p(x_n) \\
\approx p^{np}(1 - p)^{n(1-p)}
\]

Taking the logarithm

\[
\log[p(x_1, \ldots, x_n)] \approx -n [p \log(p) + (1 - p) \log(1 - p)] \\
= nH(X)
\]

\[p(x_1, \ldots, x_n) \approx 2^{-nH(X)}\]

there are at most $2^{nH(X)}$ typical sequences.

the typical sequences can be coded with $nH(X)$ bits.
Formalization

Take the $2^{nH(X)}$ most probable sequences to be the typical ones.

The probability that a sequence is typical goes to unity when $n \to \infty$.

⇒ The scheme is said to be reliable.

**Shannons noiseless channel coding theorem**

For a i.i.d source with entropy $H(X)$ there exist a reliable compressing scheme with rate $H(X)$.

For binary information, $n$ bits can reliably be compressed to $nH(X)$ bits when $n \to \infty$. 
Practical coding procedure

1) Take entire source information, cut into pieces of length $n$
2) Go through each piece.
   - If the piece is typical, code it with one of $nH(X)$ bits.
   - If the piece is atypical, abort.
3) Transmit/recieve the codes.
4) Decode/decompress the information.
5) Retrieve original information.

For large enough $n$ this scheme has a high success rate.

More refined schemes exist.
Quantum information

**Question:** How much can quantum information be compressed, transmitted and then decompressed without losses?

As a model for an information source we take a sequence of identical density matrices $\rho$ in the Hilbert space $H$

We want to transmit $\rho \otimes \rho \otimes \ldots \equiv \rho^{\otimes n}$

- The spectral decomposition $\rho = \sum_x \lambda_x |x\rangle \langle x|$ $\Rightarrow$ we have $S(\rho) = H(X)$ with $p_x = \lambda_x$.

- We can define a *typical subspace* spanned by the $2^{nS(\rho)}$ states $|x_1 \rangle |x_2 \rangle \ldots |x_n \rangle$ which have the largest probabilities $p(x_1, \ldots, x_n)$

- Project $\rho^{\otimes n}$ onto the typical subspace. The projection operator is
  $$P(n) = \sum_{x \in typical} |x_1 \rangle \langle x_1| \otimes \ldots \otimes |x_n \rangle \langle x_n|$$

  This is a quantum operation $C^n$
- Transmit the projected state $P(n) \rho \otimes^n P(n)$
- Transform the projected state to $\approx \rho \otimes^n$. This is a quantum operation $\mathcal{D}^n$

Success of the protocol is measured in terms of the fidelity

$$F(\rho, \mathcal{D}^n \circ \mathcal{C}^n(\rho))$$

where $\mathcal{D}^n \circ \mathcal{C}^n$ is the total quantum operation and $F(\rho, \sigma) = \text{tr} \left[ \rho^{1/2} \sigma \rho^{1/2} \right]$

It can be shown that for $n \to \infty$ we have $F \to 1$

---

Schumacher's noiseless quantum channel coding theorem

For a source emitting identical density matrices $\rho$ with entropy $S(\rho)$ there exist a reliable compressing scheme with rate $S(\rho)$

For qubits, $n$ qubits can realiably be compressed to $nS(\rho)$ bits when $n \to \infty$. 
12.3 Classical information over noisy channels

Classical channel

**Example:** Binary channel
Bit flips makes the transmission imperfect, noisy classical channel

\[
\begin{array}{c}
0 \quad 1 - p \\
\downarrow \quad p \\
\downarrow \quad p \\
1 - p \quad 0 \\
\end{array}
\]

Probability \( p \) to flip during transmission

We have the channel \( \mathcal{N} \)

If 500 of transmitted 1000 bits can be recovered: rate \( R = \frac{500}{1000} = \frac{1}{2} \).

The maximum achievable rate of the channel \( \mathcal{N} \) is the **channel capacity**

**Problem:** What is the channel capacity \( C(\mathcal{N}) \)?
Random coding (proof of principle)

Let's assume Alice want to transfer information at a rate $R$ through the channel to Bob.

- Construct a code book with $nR$ code words $x^j$ of length $n$. For each word sample $n$ bits randomly with probability $q$ for 0 and $1 - q$ for 1, e.g. $q = 1/2$

  $x^1 = 00101001101010100101100100100010001001110011001101001010011010110010010100101101011010010110100100010011100010011000100$

  $x^2 = 1101000101001101001101000010100100100101001100100101001011010110101101001011010010010011100010010100101101010010110100100010011100010011000100$

- Sending the words through the channel typically $np$ bits are flipped, eg

  $x^1 = 001\overline{0}1001\overline{1}0110101100\overline{1}1001\overline{0}0010\overline{0}00110110011001\overline{0}0011000100111000110010010100101101011010010110100100010011100010011000100$

  $y^1 = 0011110001101110101101001001001000100110001001100010011000100$

  The Hamming distance between $x^1$ and $y^1$ is then $np$

  Hamming distance is number of bits differing
The possible outputs $y^1$ (due to different bit flips) are inside a Hamming sphere in code word space with radius $np$.

- There are $2^{nH(p)}$ typical outputs $y^1$, since there are
  $$\binom{n}{np} \approx 2^{nH(p)}$$
  different outputs resulting from $np$ bit flips.

  $\Rightarrow$ the Hamming sphere typically contains $2^{nH(p)}$ elements.

- The same holds for all $nR$ code words $x^j$. As long as all $nR$ Hamming spheres are separated in space Bob can identify each received code word $y^j$ as the sent code word $x^j$ and completely reconstruct the information sent by Alice.
The scheme becomes unreliable when the spheres start to overlap.

If we let the bit $Y_k$ in a received code word be the result of sending the bit $X_k$, we have the received code word $y = \{Y_1, \ldots, Y_n\}$

$$\Rightarrow$$ there are $2^{nH(Y)}$ typical received code words.

The overlap of Hamming spheres starts to be considerable when the number of elements in all Hamming spheres are of order of the number of typical received words $\Rightarrow$

$$2^{nH(p)} \times 2^{nR} < 2^{nH(Y)} \Rightarrow R < H(Y) - H(p)$$

is the condition on the rate $R$ for a reliable transmission.
The rate is maximized by choosing for the input distribution $q = 1/2$

$$\Rightarrow H(X) = 1 \quad \text{and thus} \quad H(Y) = 1$$

i.e. bit flips have no effect on the entropy

$$\Rightarrow$$

The channel capacity is

$$C(\mathcal{N}) = 1 - H(p)$$

---

**Shannons noisy channel coding theorem (no proof)**

For a noisy channel $\mathcal{N}$ the capacity is given by

$$C(\mathcal{N}) = \max_{p(x)} H(X : Y)$$

where $p(x)$ is the distribution of $X$ and $Y$ the output random variable.

---

For the binary case, the random coding model

$$H(X : Y) = H(Y) - H(X|Y) = H(Y) - H(p)$$

$$\Rightarrow C(\mathcal{N}) = 1 - H(p) \quad \text{Maximial capacity}$$
Compressing and coding in noisy channel

- For a binary information source with entropy $H(X)$ one can without losses compress $n$ bits of information to $nH(X)$ bits.

- To reliably send $nH(X)$ bits over a noisy channel with capacity $C(N) = 1 - H(p)$ requires

  $$n \frac{H(X)}{1 - H(p)}$$

  bits. This is an optimal scheme.
Let’s assume Alice wants to transfer information at a rate $R$ through the noisy quantum channel, characterized by the operation $\mathcal{E}$, to Bob.

- Construct a code book with $nR$ code words $x^j$ of length $n$, just as for the classical case.
- Translate each classical code word to a quantum state by choosing for bits 0 a qubit density matrix $\rho_0$ and bit 1 a qubit density matrix $\rho_1$
  \[ x^j = \{x_1, x_2, \ldots, x_n\} \rightarrow \rho^j = \rho x_1 \otimes \rho x_2 \otimes \ldots \otimes \rho x_n \]
- When sending $\rho^j$, Bob will receive the state $\sigma^j = \mathcal{E}^{\otimes n}(\rho^j)$
- Bob then performs a POVM measurement with elements $\{E_j\}$, where each element $E_j$ corresponds to a given code word.

**Problem:** What is the product state capacity $C^{(1)}(\mathcal{E})$ of the channel?
Holevo-Schumacher-Westmoreland theorem

Let $\mathcal{E}$ be a trace preserving operation. The product state capacity is

$$C^{(1)}(\mathcal{E}) = \max_{\{p_j, \rho_j\}} \left\{ S \left[ \mathcal{E} \left( \sum_j p_j \rho_j \right) \right] - \sum_j p_j S \left[ \mathcal{E}(\rho_j) \right] \right\}$$

where the maximum is over all ensembles $\{p_j, \rho_j\}$. Here no proof.

12.4 Quantum information over noisy channels

Still an open problem. Here no discussion.
12.5 Entanglement

Focus on entanglement characterization, not in book (rapidly developing field). Small part on entanglement as resource.

Pure state entanglement

Ex: Bipartite entanglement.

A quantum state $|\psi_{AB}\rangle$ in $\mathcal{H}_A \otimes \mathcal{H}_B$ is entangled if

$$|\psi_{AB}\rangle \neq |\psi_A\rangle \otimes |\psi_B\rangle$$

for any states $|\psi_A\rangle$ in $\mathcal{H}_A$ and $|\psi_B\rangle$ in $\mathcal{H}_B$.

**Question**: How large is the entanglement?

The entanglement $E(|\psi_{AB}\rangle)$ is given by the von Neumann entropy of the reduced density matrix

$$E(|\psi_{AB}\rangle) = S(\rho_A) = S(\rho_B)$$

where $\rho_A = \text{tr}_B(|\psi_{AB}\rangle\langle\psi_{AB}|)$ and $\rho_A = \text{tr}_B(|\psi_{AB}\rangle\langle\psi_{AB}|)$.

Note that $S(\rho_A) = S(\rho_B)$ follows from the Schmidt decomposition

$$|\psi_{AB}\rangle = \sum_i \lambda_i |i_A\rangle|i_B\rangle$$
**Example:** Bell state

**Derivation**

Extendable to multipartite pure entanglement.

**Two qubit case**

One can write the state

$$|\psi_{AB}\rangle = c_{00}|00\rangle + c_{01}|01\rangle + c_{10}|10\rangle + c_{11}|11\rangle$$

The entanglement can be written

$$E(|\psi_{AB}\rangle) = H_{bin}(q)$$

$$q = \frac{1 + \sqrt{1 - 2C^2}}{2}$$

where $C$ is the concurrence

$$C = 2\sqrt{\det(cc^\dagger)}$$

$$c = \begin{pmatrix} c_{00} & c_{01} \\ c_{10} & c_{11} \end{pmatrix}$$

The concurrence $0 \leq C \leq 1$ is used as a measure of entanglement itself.
Mixed state entanglement

Ex: Bipartite entanglement.

A quantum state $\rho_{AB}$ in $\mathcal{H}_A \otimes \mathcal{H}_B$ is entangled if

$$\rho_{AB} \neq \sum_i p_i |\psi_{Ai}\rangle \langle \psi_{Ai}| \otimes |\psi_{Bi}\rangle \langle \psi_{Bi}| \quad \sum_i p_i = 1$$

for any states $|\psi_{Ai}\rangle$ in $\mathcal{H}_A$ and $|\psi_{Bi}\rangle$ in $\mathcal{H}_B$.

There is no unique measure of mixed state entanglement. An important quantity is entanglement entropy (entanglement of formation)

$$E(\rho_{AB}) = \min_{\{p_i, |\psi_{ABi}\rangle\}} \sum_i p_i E(|\psi_{ABi}\rangle)$$

where $|\psi_{ABi}\rangle = |\psi_{Ai}\rangle \otimes |\psi_{Bi}\rangle$

The entanglement entropy forms an upper bound for

- Entanglement cost/dilution $|\psi_{Bell}\rangle \rightarrow \rho$
- Entanglement of distillation $\rho \rightarrow |\psi_{Bell}\rangle$
Two qubit case
For two qubits there is a closed expression

\[ E(\rho_{AB}) = H_{bin}(q) \quad q = \frac{1 + \sqrt{1 - 2C^2}}{2} \]

where \( C \) is the concurrence

\[ C = \max \left\{ \sqrt{\lambda_1} - \sqrt{\lambda_2} - \sqrt{\lambda_3} - \sqrt{\lambda_4} \right\} \]

with \( \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4 \) the eigenvalues of

\[ \rho_{AB}(\sigma_y \otimes \sigma_y)\rho^*_{AB}(\sigma_y \otimes \sigma_y) \]

**Example:** Werner state
The Werner state is

\[ \rho_W = \frac{1}{4}(1 - \xi)I + \xi|\psi_S\rangle\langle\psi_S| \quad |\psi_S\rangle = \frac{1}{\sqrt{2}}[|01\rangle - |10\rangle] \]

with

\[ C = \max \left\{ \frac{1}{2}(3\xi - 1), 0 \right\} \]

finite for \( \xi > 1/3 \).
Note: Bell inequality is violated for \( \xi \leq 1/\sqrt{2} \).
12.6 Quantum key distribution

Old problem: how to transfer information between Alice and Bob in a secure way? The problem is typically to transfer the coding key.

Classical key distribution
Two different approaches

- Private key distribution
  i) At an earlier point Alice and Bob meet and decide about a secret key
  ii) Alice decodes the message with the key.
  iii) Alice sends the coded message to Bob.
  iv) Bob decodes the message

Pros and cons

+ For a key as long as the message the approach is provably safe (Vernam cipher)
- The key can only be used once to preserve complete safety.
- The distribution of the key can be as difficult as distributing the secret message itself.
- If the spy Eve gets her hands on the key she can decode the message.
Public key distribution (no details)

i) Alice announces a public key ⇒ anyone can code a message with the key and send her.
ii) Bob uses Alice's key and codes his message to her.
iii) Bob sends Alice the message.
iv) Alice receives the message and decodes it. Importantly, the public key is such that only Alice can decode the message.

Such schemes were invented only in the seventies. The most spread is RSA which is used in many situations on the internet.

Pros and cons

+ Alice and Bob do not have to meet to exchange keys.
+ If Eve reads the coded message it is almost impossible to decode it.
- Advanced schemes, like RSA, are based on the difficulty to factorize large numbers into prime factors ⇒ can be broken by a quantum computer.
Quantum key distribution

We consider a noiseless channel. There are a number of different protocols for quantum key distribution, here we consider BB84.

i) Alice chooses randomly two bit strings $a, b$ of length $n$, e.g.

$$a = 001010011010110100110101100101001010011010110010100101001011$$

$$b = 1101001010011010000101001001010010110101101001011$$

ii) Alice encodes each pair of bits $a_k, b_k$ into a qubit as

$$|\psi_{00}\rangle = |0\rangle$$
$$|\psi_{10}\rangle = |1\rangle$$
$$|\psi_{01}\rangle = \frac{1}{\sqrt{2}} [|0\rangle + |1\rangle]$$
$$|\psi_{11}\rangle = \frac{1}{\sqrt{2}} [|0\rangle - |1\rangle]$$

$\Rightarrow$$\begin{cases} 
\text{for } b_k = 0 \text{ she codes } a_k \text{ in the Z-basis} \\
\text{for } b_k = 1 \text{ she codes } a_k \text{ in the X-basis}
\end{cases}$
iii) Alice sends the message to Bob.
iv) Bob measures the qubits choosing randomly the basis Z or X. He stores his basis choice in $b'$ and his measurement outcomes in $a'$.
v) Alice and Bob announce and compare $b$ and $b'$.
vi) Alice and Bob discard the bits where they measured differently, i.e. $b_k \neq b'_k$. This will typically be half of the bits, $n/2$. The rest of the bits, $a_k$ and $a'_k$, for which $b_k = b'_k$ are kept.
vii) Alice selects a subset $m$ of the $n/2$ remaining bits and announce the preparation $a_k$. Bob compares Alice's preparation to his own measurement result.
   - If they agree, i.e. $a_k = a'_k$, they decide to use the remaining $n/2 - m$ bits as a key.
   - If a lot of the bits do not agree they can suspect that Eve has tried to measure and then they abort.

**Important:** since Eve does not know what basis Alice chooses to prepare the state in for each bit she will typically measure in a different basis half of the times. This inevitably will modify the bits Eve measures, which Alice and Bob will notice.

Quantum key distribution can in principle be made safe