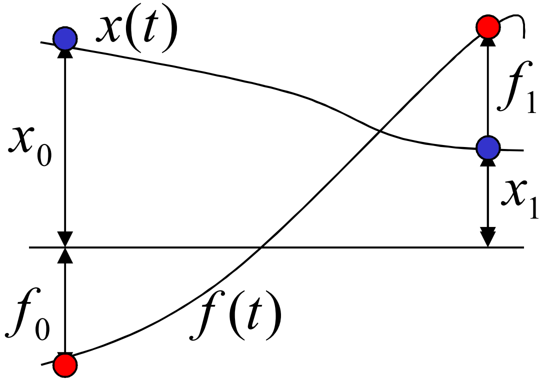
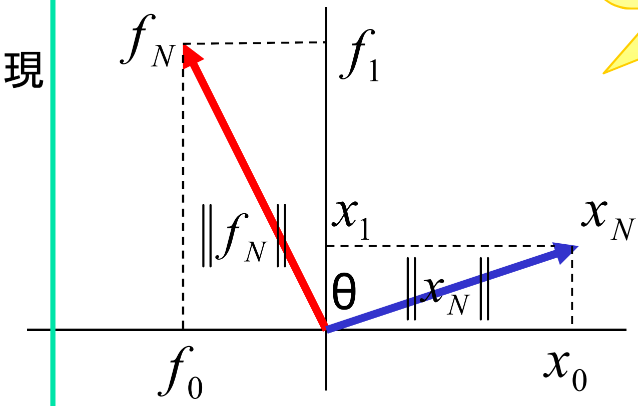
相関係数

復習：
信号(データ)をベクトルで表現

元が連続信号なら
離散化してベクトルで表現



とりあえず
ここでは2次元で
以下のようなベクトルがあるとして



$$x_N = \begin{pmatrix} x_0 \\ x_1 \end{pmatrix}$$

$$f_N = \begin{pmatrix} f_0 \\ f_1 \end{pmatrix}$$

２つの信号(データ)の関係をみたい場合

どうする？

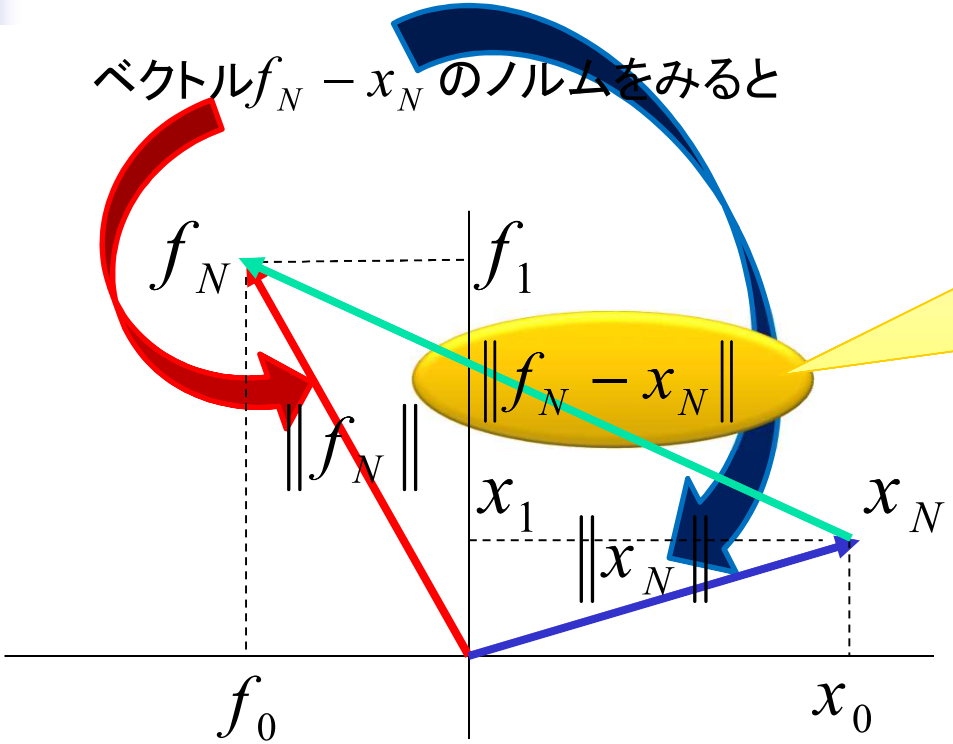
２つの信号ベクトルの

・<u>距離</u>をみる？

・<u>角度</u>をみる？

ベクトルの距離で考えてみる

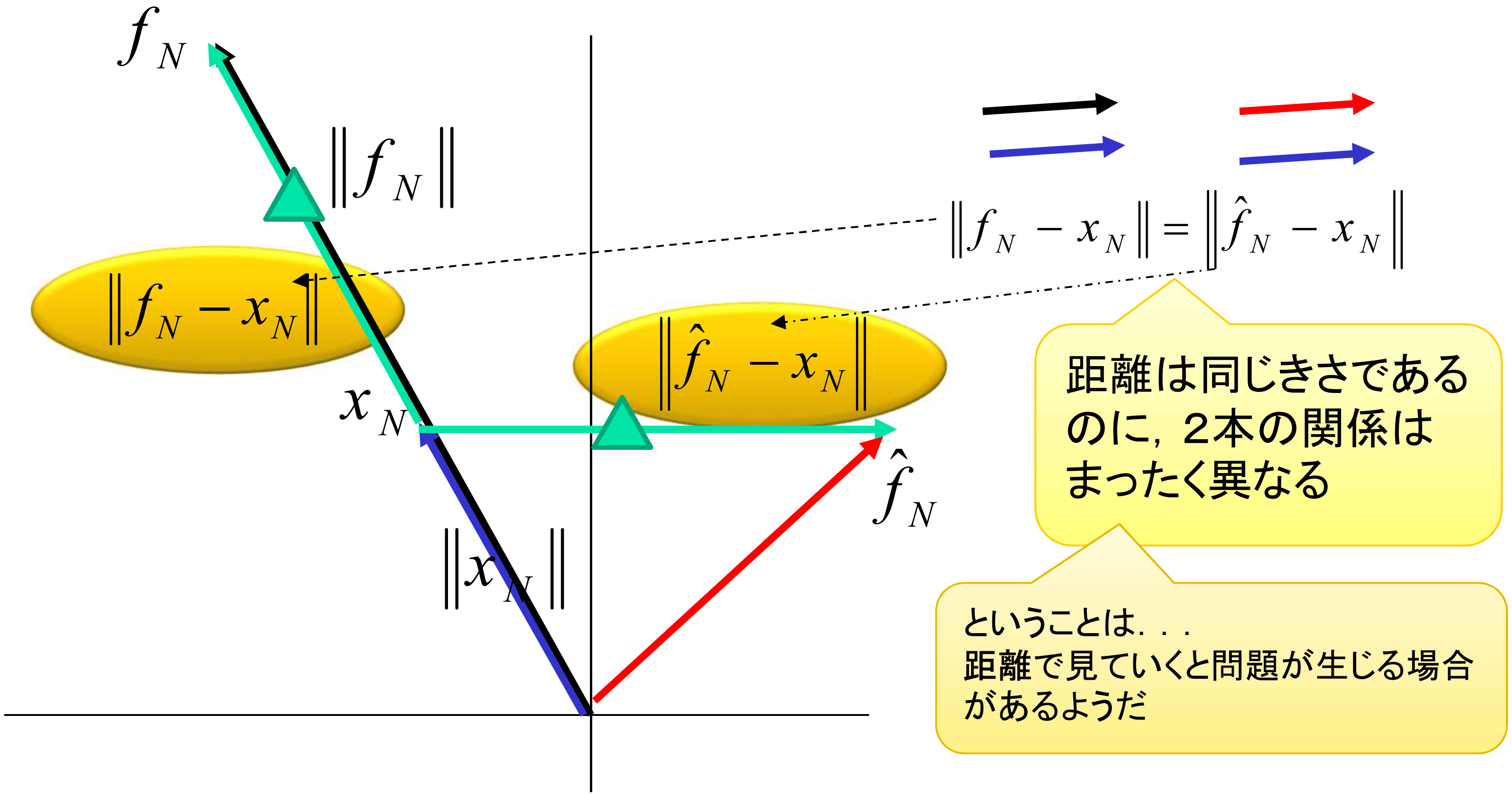
ベクトル$f_N - x_N$のノルムをみると



この距離の大小で
２本のベクトルの
「関係の強さ」を
見ることが出来るようだ

ベクトルの距離で考えてみる

一方，別の場合で2本のベクトルの大きさを見てみる．
すると．．．．．



相関係数計算準備（2つの信号の角度を見る）

相関係数の考え方

2つの信号の関係の強さを，「角度に基づいて」みる
信号の類似性を見ることが出来る

$$\cos\theta \longrightarrow \mathrm{r} = \frac{\langle x_N, f_N \rangle}{\|x_N\| \cdot \|f_N\|}$$

負(-)の相関も
あり得る

-1～0～1で
関係の大きさを示す

プログラムでは
これを実現する

r=0のとき
無相関（関係がない）
このとき2本のベクトルは
直交関係

内積！

$$\frac{x_N}{\|x_N\|} \qquad \frac{f_N}{\|f_N\|}$$

大きさを1に

大きさを1に

・rが0に近いデータ同士の関係が弱い（小さい）
・rが1に近いほど正の相関が強い（大きい）
・rが-1に近いほど負の相関が強い（大きい）

相関係数の計算
平均値(直流成分)を除去

相関係数をみるときの「基準」が違ってしまうと，相関係数自体の数値も異なることがある
従って，相関係数は平均値(直流成分)を取り除いた形で求められる.

$$\bar{x} = \frac{1}{N}\sum_{i=0}^{N-1} x_i, \quad \bar{f} = \frac{1}{N}\sum_{i=0}^{N-1} f_i$$

それぞれのデータについて平均値を計算し

$$x'_N = (x_0-\bar{x}, x_1-\bar{x}, \cdots, x_{N-1}-\bar{x})$$

$$f'_N = (f_0-\bar{f}, f_1-\bar{f}, \cdots, f_{N-1}-\bar{f})$$

係数を計算する前に，あらかじめ平均値を引いておく

プログラムではこれも実現する

平均を除去しておけば，単位のスケールが異なろうが，物理量が違おうが，
関係なく，その関係を求めることが出来る.
平均値の削除でどの程度相関係数に違いが出るかは，課題で実際に体感してみよう

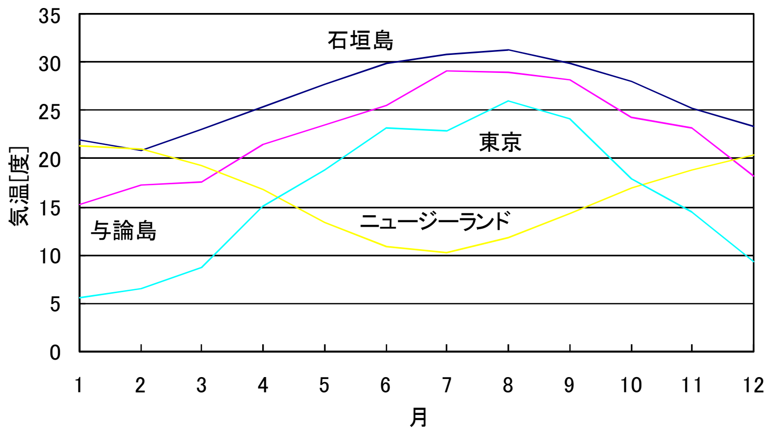
たとえば

自然現象では，気温，水温，降雨量など相関係数を利用することにより，
その関連性を数値として見ることが出来る．

気温で相関を予想してみると
　１．東京とその他日本の各都市との相関係数は似ているだろう
　２．東京と離れた都市との相関係数は１．に比べ似ていないだろう．
　　　南半球ならマイナスの相関もありそうだ
　（１．の方が気温変化が似ているので，よりr=1に近いと思われが．．．）

実際，相関係数は
　東京・ニュージーランド：-0.934
　東京・石垣：0.986
　東京・与論：0.975

もし平均を除去しなければ
　東京・ニュージーランド：0.805
　東京・石垣：0.961
　東京・与論：0.976



正の相関と負の相関

例えば「テストの点」は
・前日の勉強時間
と強い（正の）相関

・前日の食事時間
とはほぼ無相関

・前日のゲーム時間
と強い（負の）相関

の関係があると
考えられる