

データサイエンス概論I & II データサイエンス総論I & II

線形代数に基づくデータ解析の基礎

九州大学 数理・データサイエンス教育研究センター

このセクションで言いたいこと

- 各データ（ベクトル）を「いくつかのベクトル」の成分に**分解**することで、そのデータの性質がよくわかるかも
 - カレーを「肉・ジャガイモ…」成分に分解すれば、どんなカレーかわかる
- 分解は「**内積**」一発
 - 内積＝「掛け算」と「足し算」するだけ（小学生でもできる!）
- 「いくつかのベクトル」として何を用いるかは、あなた次第！
 - それによって分解結果は異なる
 - 要するに、この「いくつかのベクトル」が「データの見方」を決める
 - 例えばカレーを、「肉・ジャガイモ…」に分解するか、「ビタミンA, たんぱく質…」に分解するか、「窒素, 炭素, 酸素…」に分解するか…



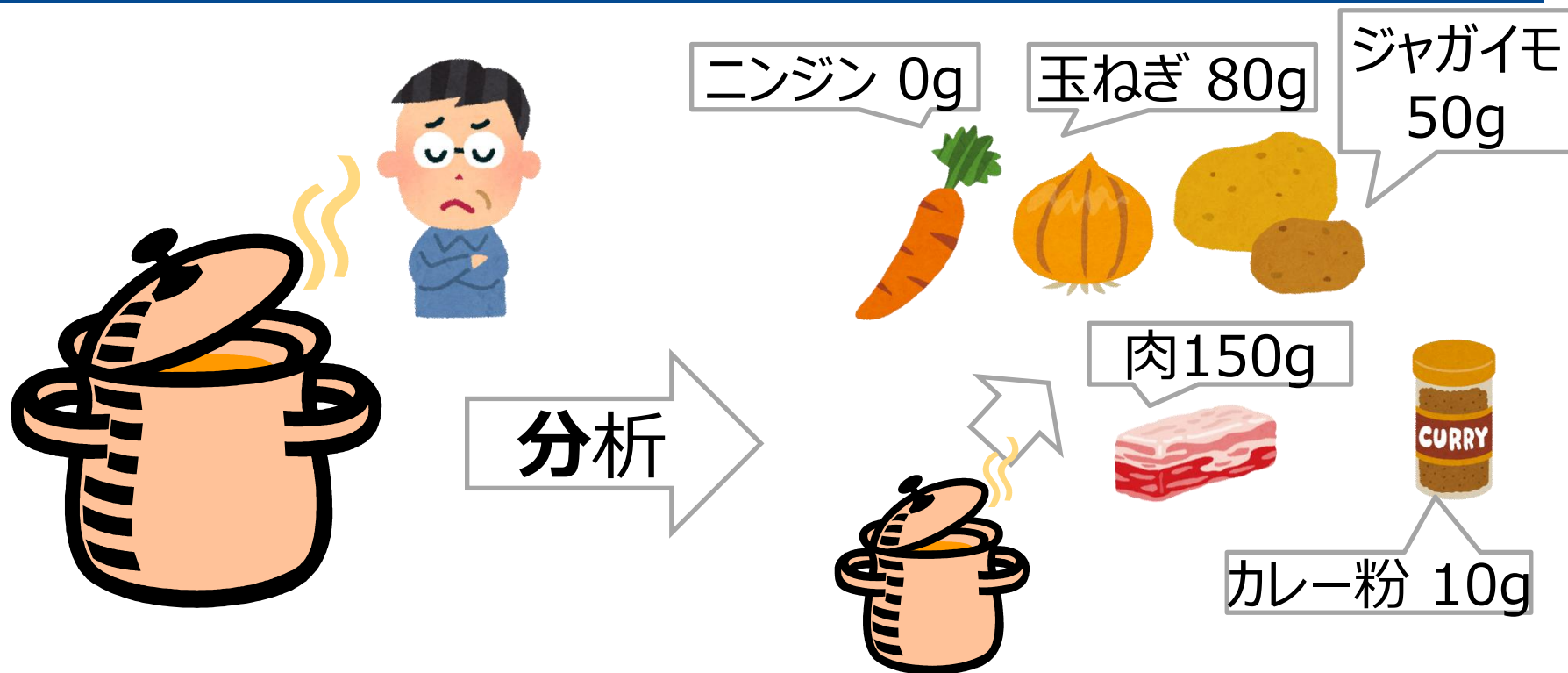
「分析」の基本的な考え方

あなたのカレーにジャガイモは何グラム入ってる？

「分析」とは？（デジタル大辞泉より）

1. 複雑な事柄を一つ一つの要素や成分に分け、その構成などを明らかにすること。「情勢の一があまい」「事故の原因を一する」
2. 哲学で、複雑な現象・概念などを、それを構成している要素に分けて解明すること。⇔総合。
3. 物質の組成を調べ、その成分の種類や量の割合を明らかにすること。

分析の例： カレーにジャガイモは何グラム入ってる？



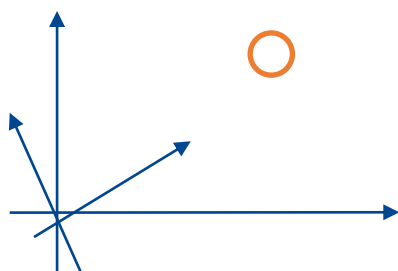
色々なものが混ざっているので
パッと見ただけでは
どんなカレーかわからない

何がどれくらい混ざっているか
わかったら、どんなカレーか
クリアになる！

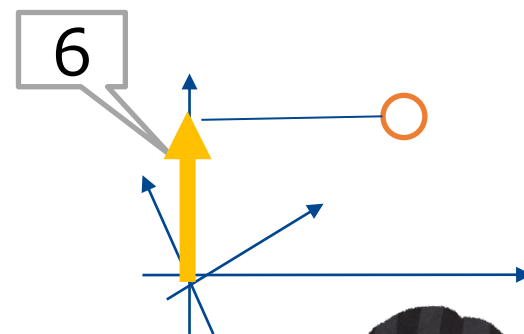
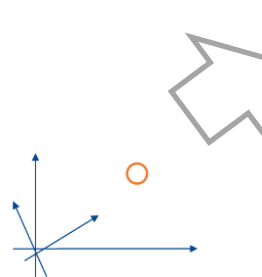
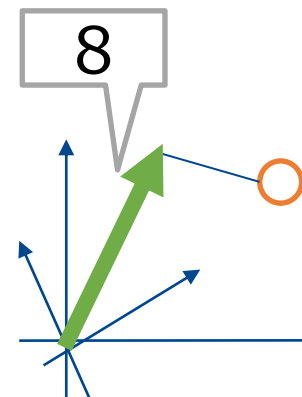
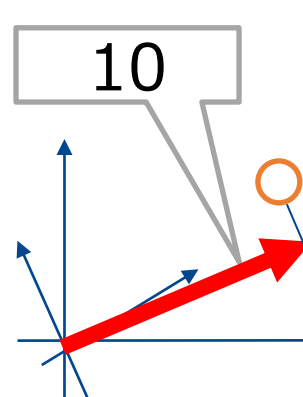


これからの話の「予告編」

カレーからベクトルへ



分析



色々なものが混ざっているので
パッと見ただけでは
どんな高次元ベクトルかわからない

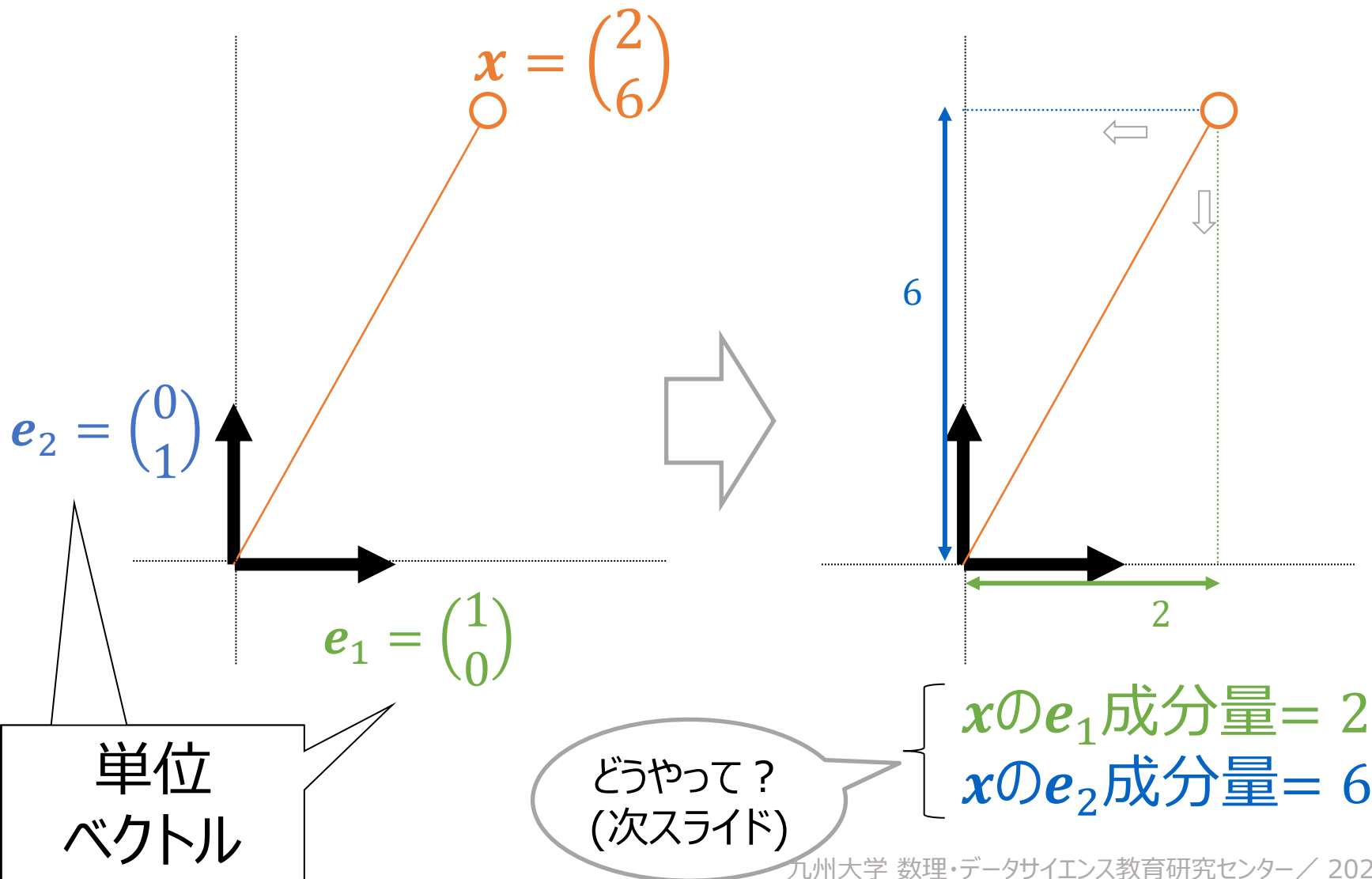
何がどれくらい混ざっているか
わかったら、どんな高次元
ベクトルかクリアになる！



ベクトルの分解と合成

高校時代に習った「内積」と「カレー」の意外な関係

ベクトルの分解



驚きの事実？ 「内積」を使えば成分量が計れる！

$x = \begin{pmatrix} 2 \\ 6 \end{pmatrix}$ の中に $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ はどれくらい入っている？

$$x \cdot e_1 = \left[\begin{pmatrix} 2 \\ 6 \end{pmatrix} \times + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right] = 2$$

掛け算と足し算

内積で x に含まれる e_1 成分量が測れる！

思い出そう：内積

内積の書き方4種
(教科書で違ったりする)

$$\boldsymbol{x} \cdot \boldsymbol{y}$$

$$(\boldsymbol{x}, \boldsymbol{y})$$

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle$$

$$\boldsymbol{x}^T \boldsymbol{y}$$

- 前にやりましたよね？

$$\boldsymbol{x} = \begin{pmatrix} 3 \\ 5 \end{pmatrix}, \boldsymbol{y} = \begin{pmatrix} 6 \\ 1 \end{pmatrix} \text{の内積} \rightarrow \boldsymbol{x} \cdot \boldsymbol{y} = 3 \times 6 + 5 \times 1 = 23$$

- 要するに、「要素どうしの積をとって，全部足す」
 - その原理で，何次元ベクトルでも計算可能

$$\begin{pmatrix} 3 \\ 5 \end{pmatrix} \text{と} \begin{pmatrix} 6 \\ 1 \end{pmatrix} \text{の内積} \Rightarrow \left. \begin{array}{l} \begin{pmatrix} 3 \\ 5 \end{pmatrix} \times \begin{pmatrix} 6 \\ 1 \end{pmatrix} = 18 \\ \phantom{\begin{pmatrix} 3 \\ 5 \end{pmatrix}} \times \phantom{\begin{pmatrix} 6 \\ 1 \end{pmatrix}} = 5 \end{array} \right\} 18 + 5 = 23$$

$$\begin{pmatrix} 3 \\ 5 \\ 2 \end{pmatrix} \text{と} \begin{pmatrix} 6 \\ 1 \\ 2 \end{pmatrix} \text{の内積} \Rightarrow \left. \begin{array}{l} \begin{pmatrix} 3 \\ 5 \\ 2 \end{pmatrix} \times \begin{pmatrix} 6 \\ 1 \\ 2 \end{pmatrix} = 18 \\ \phantom{\begin{pmatrix} 3 \\ 5 \\ 2 \end{pmatrix}} \times \phantom{\begin{pmatrix} 6 \\ 1 \\ 2 \end{pmatrix}} = 5 \\ \phantom{\begin{pmatrix} 3 \\ 5 \\ 2 \end{pmatrix}} \times \phantom{\begin{pmatrix} 6 \\ 1 \\ 2 \end{pmatrix}} = 4 \end{array} \right\} 18 + 5 + 4 = 27$$

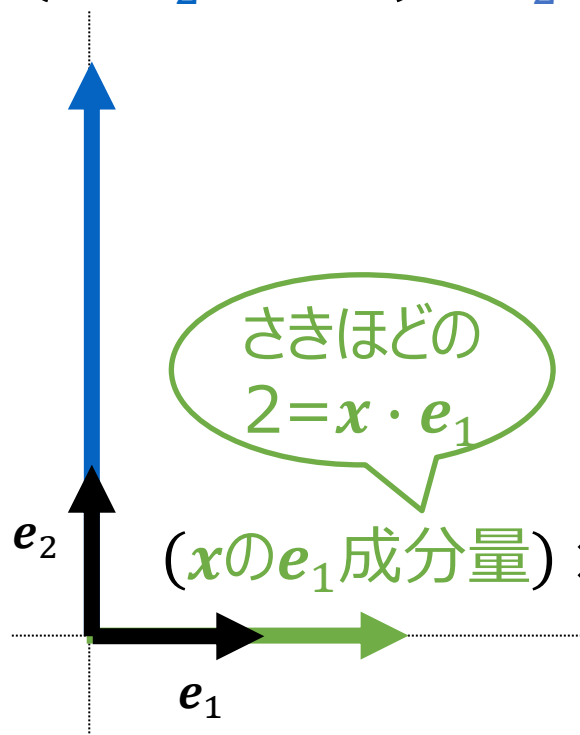
※この調子で，4次元でも，100万次元でも可能

ベクトルの合成

さきほどの
 $6 = x \cdot e_2$

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$(x \text{ の } e_2 \text{ 成分量}) \times e_2 = \begin{pmatrix} 0 \\ 6 \end{pmatrix}$$



さきほどの
 $2 = x \cdot e_1$

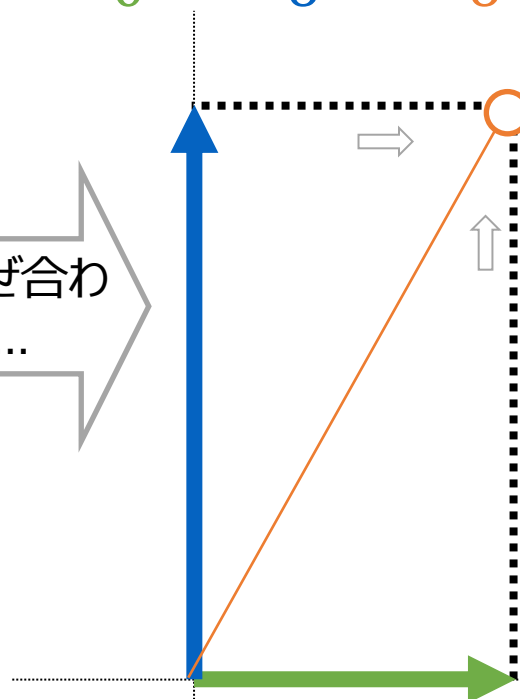
$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$(x \text{ の } e_1 \text{ 成分量}) \times e_1 = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

① 単位ベクトルを成分量どおりにして

② 混ぜ合わせると...

$$\begin{pmatrix} 2 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 6 \end{pmatrix} = \begin{pmatrix} 2 \\ 6 \end{pmatrix} = x$$

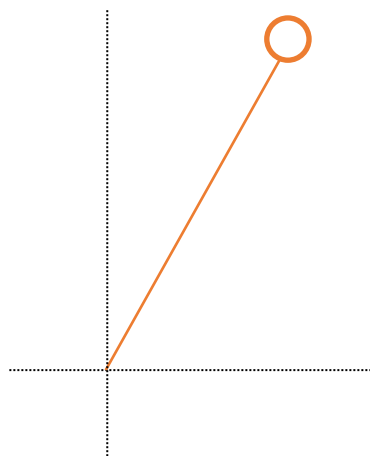


③ 元に戻る！
なんと美しい！



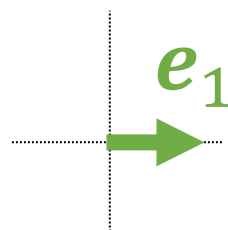
カレーとの関係： 単位ベクトルとはジャガイモである

ベクトル x



1単位

には



が

$x \cdot e_1$ だけ入っている



カレー x

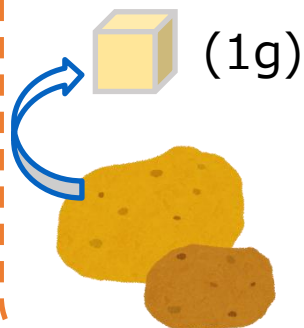
には

ジャガイモ

が

50個

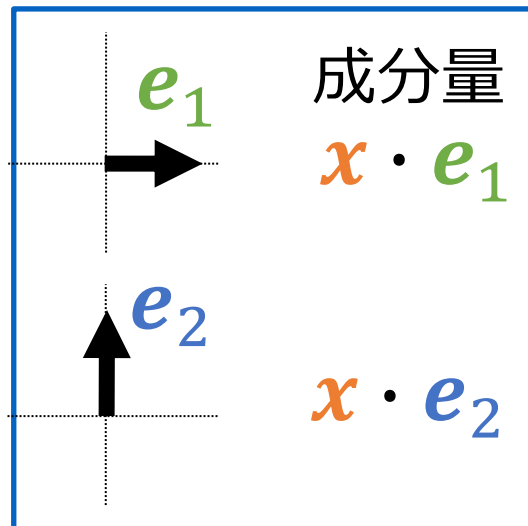
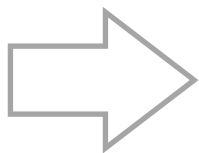
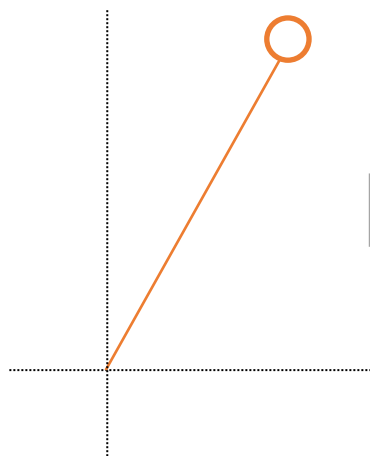
だけ入っている



要するに50g

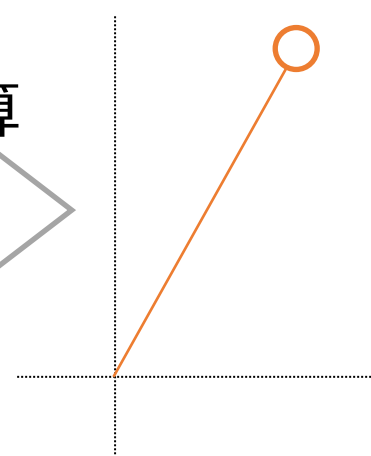
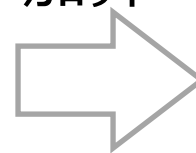
カレーとの関係： 単位ベクトルとはジャガイモやニンジンである

ベクトル x

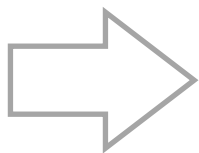


元と同じベクトル x

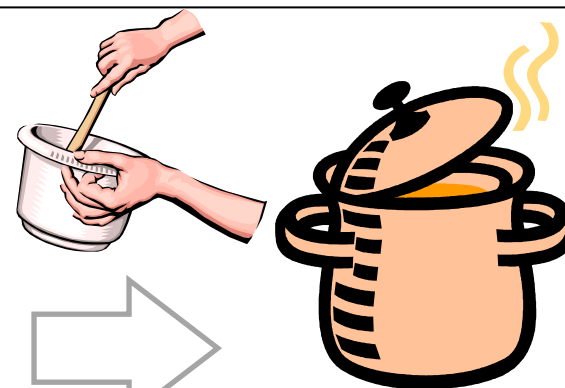
加算



カレー x

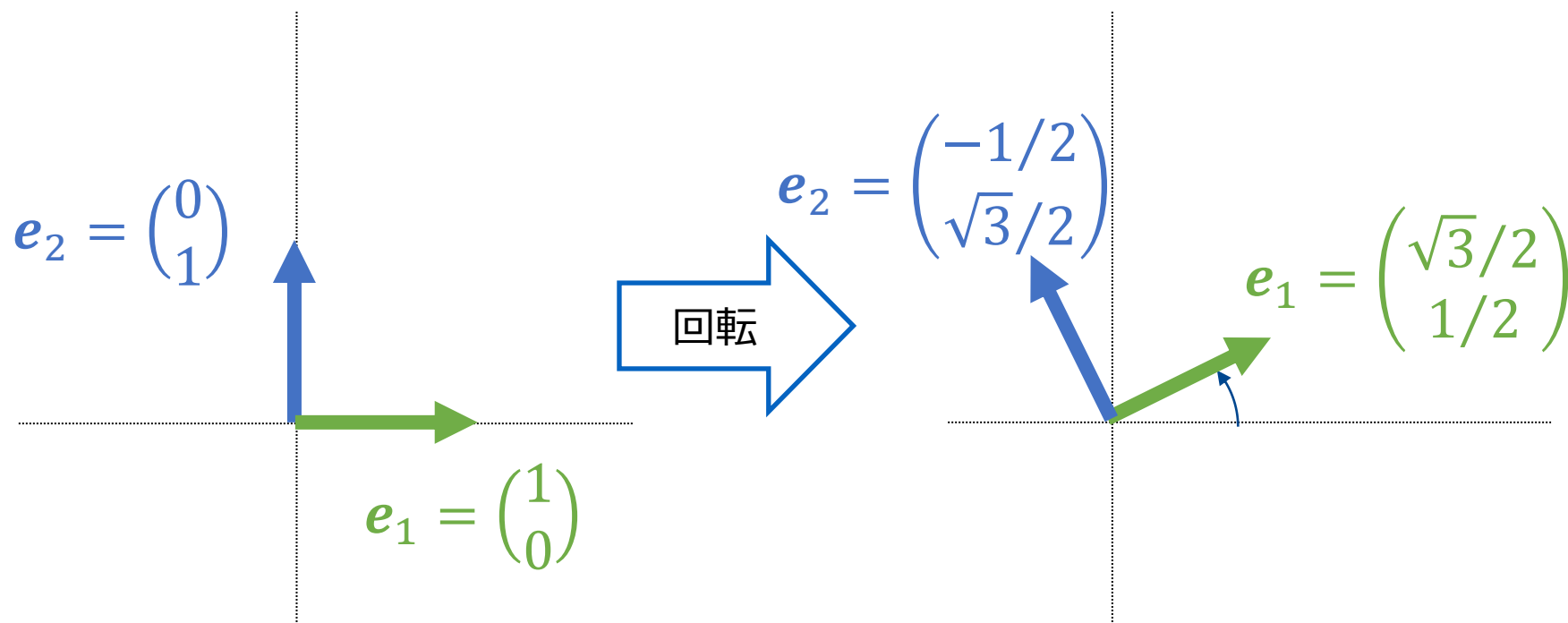


ジャガイモ ○ g
にんじん □ g
肉 △ g
:



元と同じカレー x

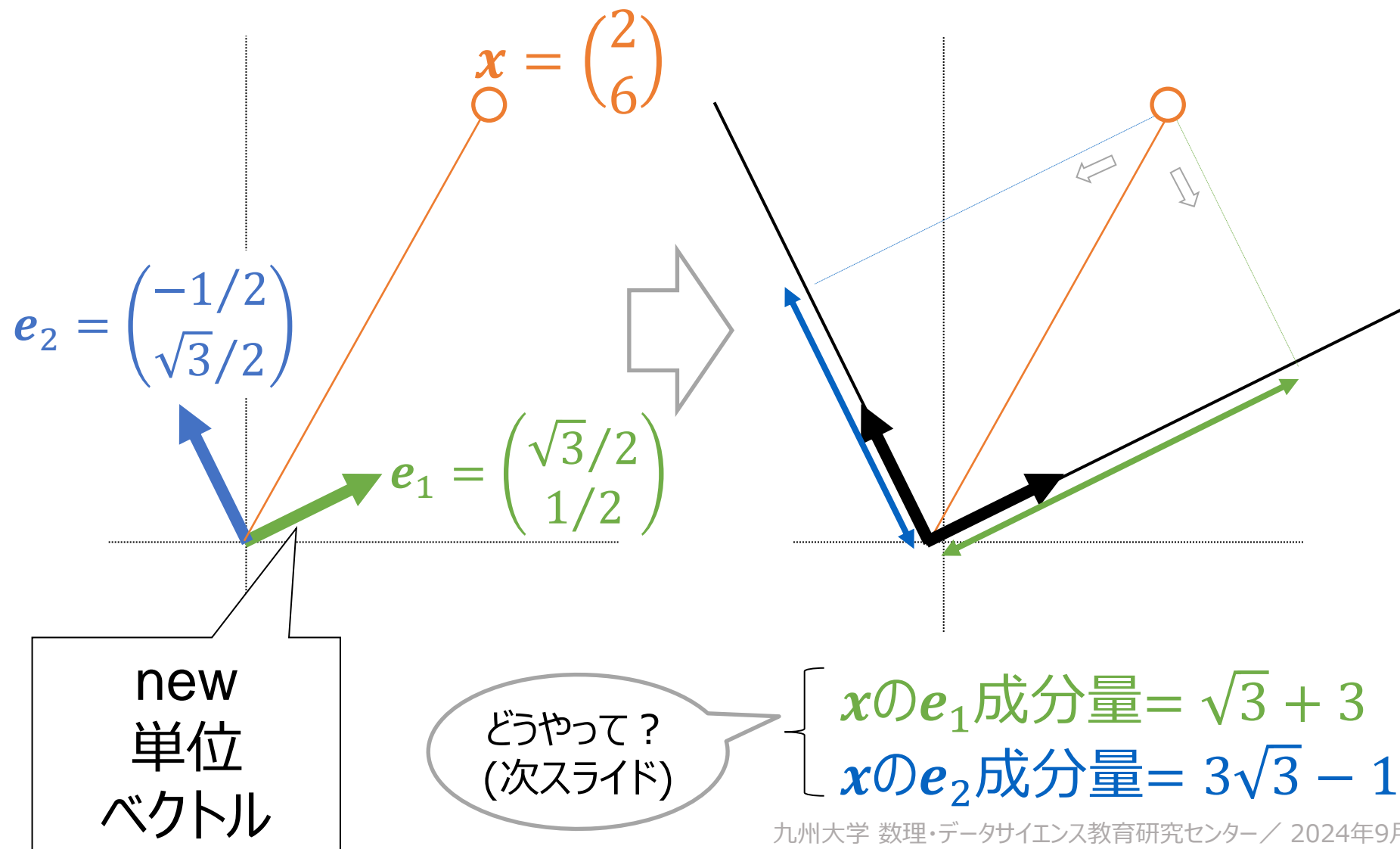
さて，単位ベクトルを回転してみる！



皆さんの良く知っている
単位ベクトル

左を30°回転させてできた
単位ベクトル

回転してできた単位ベクトルでも 先ほどのベクトルを分解できる



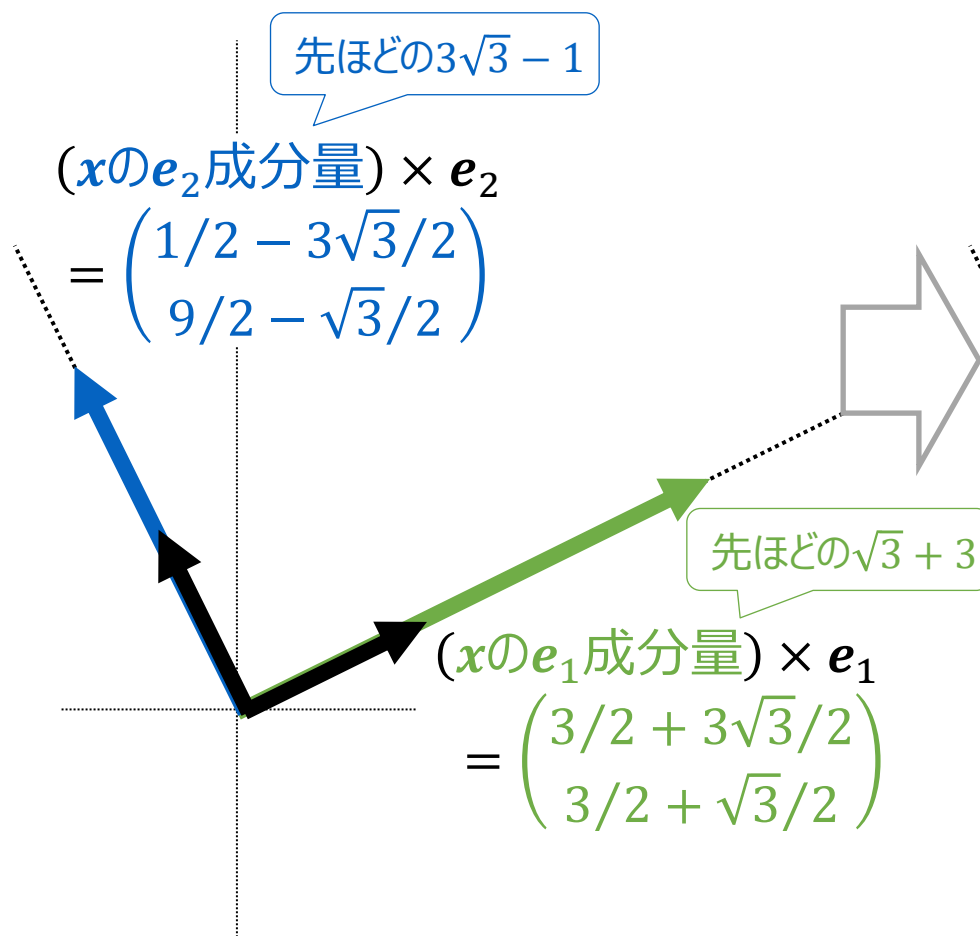
めんどくさそうな単位ベクトルでも
「内積」を使えば成分量が計れる！

$x = \begin{pmatrix} 2 \\ 6 \end{pmatrix}$ の中に $e_1 = \begin{pmatrix} \sqrt{3}/2 \\ 1/2 \end{pmatrix}$ はどれぐらい入っている？

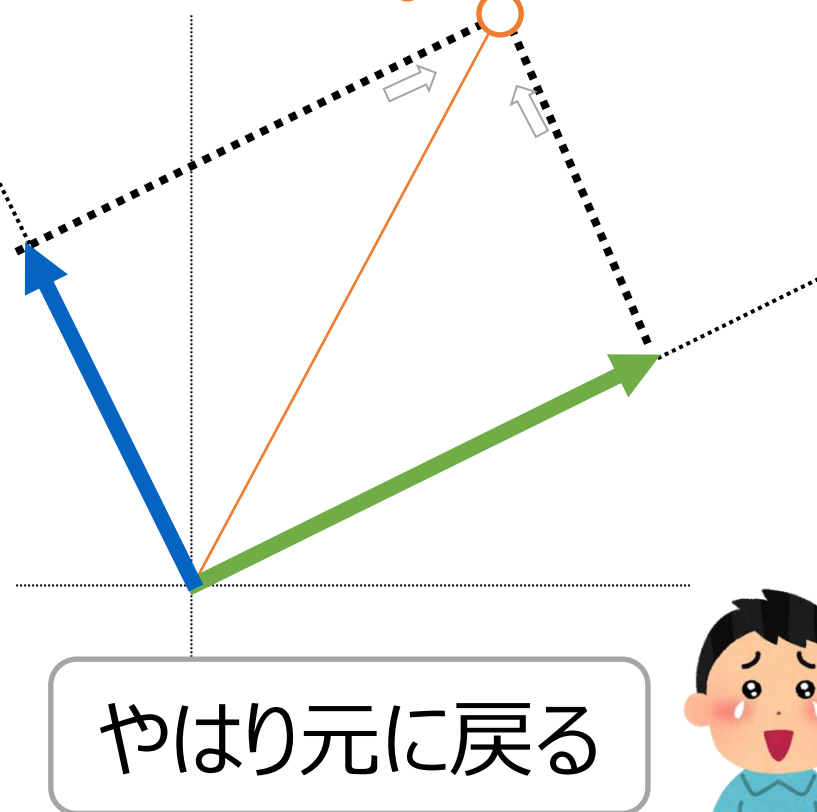


$$x \cdot e_1 = \left(\begin{pmatrix} 2 \\ 6 \end{pmatrix} \times \begin{pmatrix} \sqrt{3}/2 \\ 1/2 \end{pmatrix} \right) = \sqrt{3} + 3$$

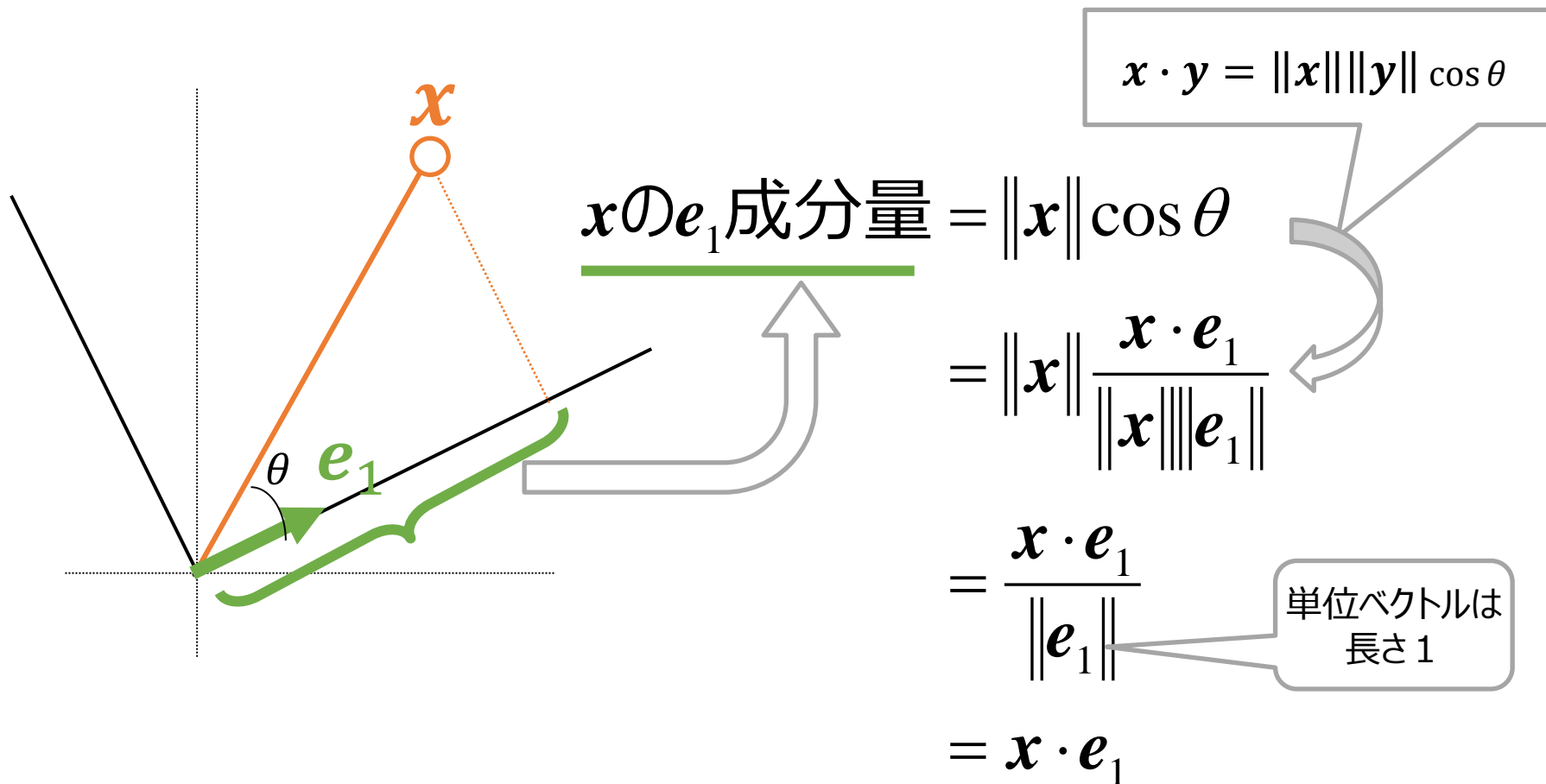
ベクトルの合成



$$\begin{pmatrix} 3/2 + 3\sqrt{3}/2 \\ 3/2 + \sqrt{3}/2 \end{pmatrix} + \begin{pmatrix} 1/2 - 3\sqrt{3}/2 \\ 9/2 - \sqrt{3}/2 \end{pmatrix} = \begin{pmatrix} 2 \\ 6 \end{pmatrix} = x$$

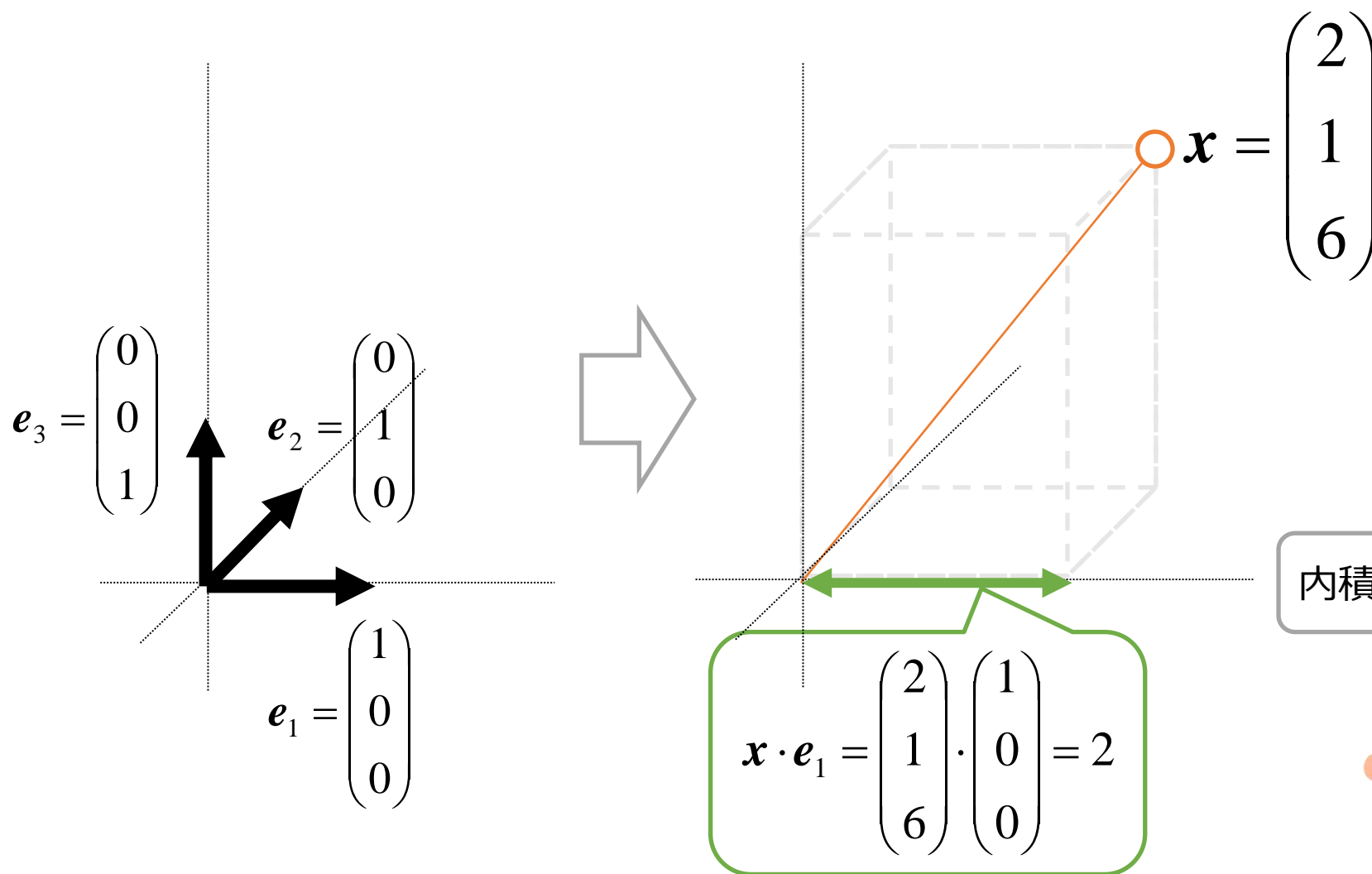


参考：内積で成分量を計っている様子を幾何的にも見てみる



- このように単位ベクトルとの内積によりその成分量が求まる

以上は2次元の場合の例でしたが、
3次元でも（もちろんそれ以上でも）同じ話



分解と合成, 一連の話をまとめると...(1/2)

- \mathbf{x} を d 次元ベクトルとする
- 内積により第 i 番目の単位ベクトル \mathbf{e}_i の成分量 α_i を出す

$$\alpha_i = \mathbf{x} \cdot \mathbf{e}_i$$

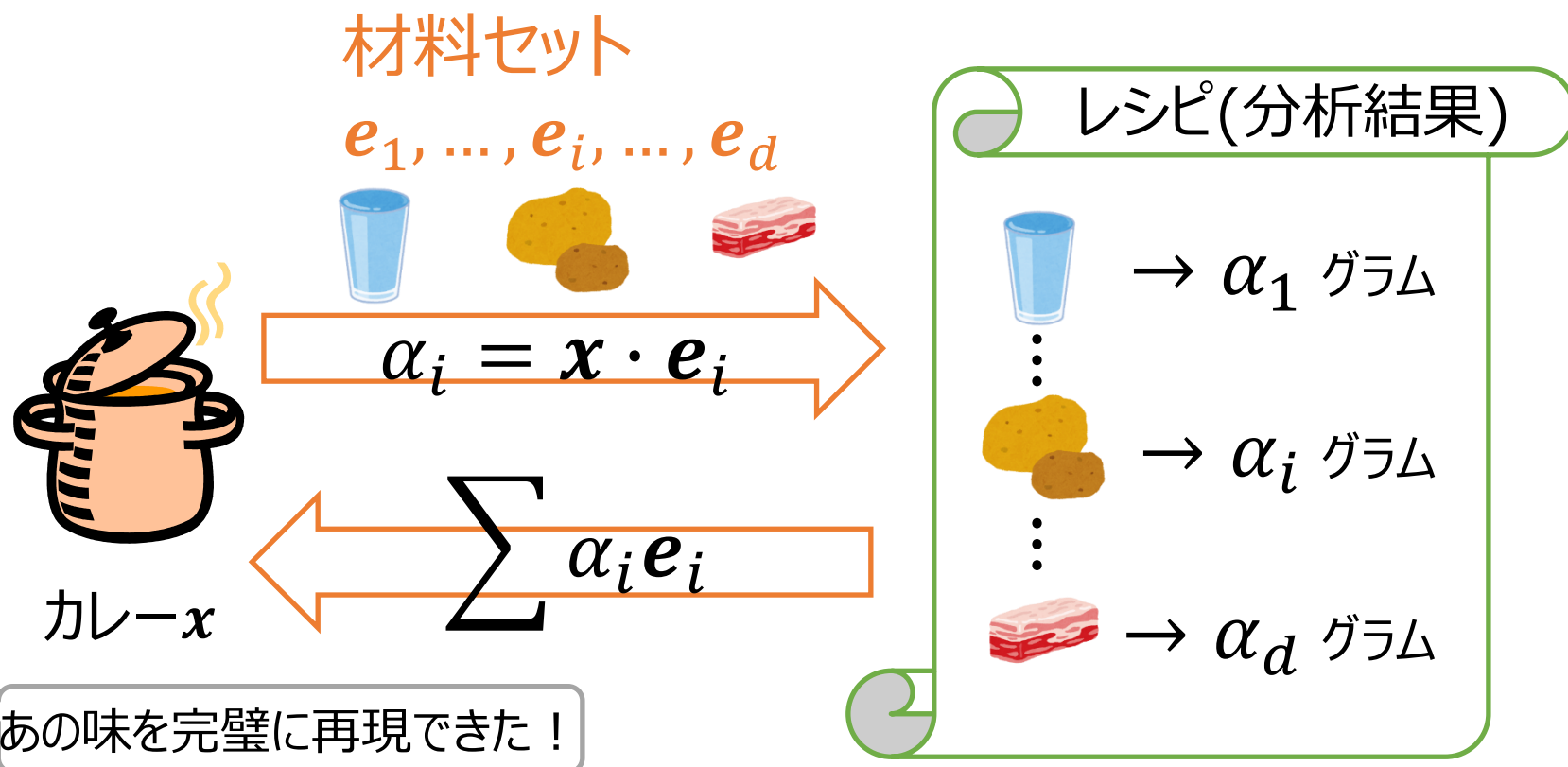
- ちなみに各 \mathbf{e}_i も d 次元ベクトル
- その成分量に応じて単位ベクトルを混ぜれば, 元の \mathbf{x} に!

$$\mathbf{x} = \alpha_1 \mathbf{e}_1 + \cdots + \alpha_i \mathbf{e}_i + \cdots + \alpha_d \mathbf{e}_d = \sum \alpha_i \mathbf{e}_i$$

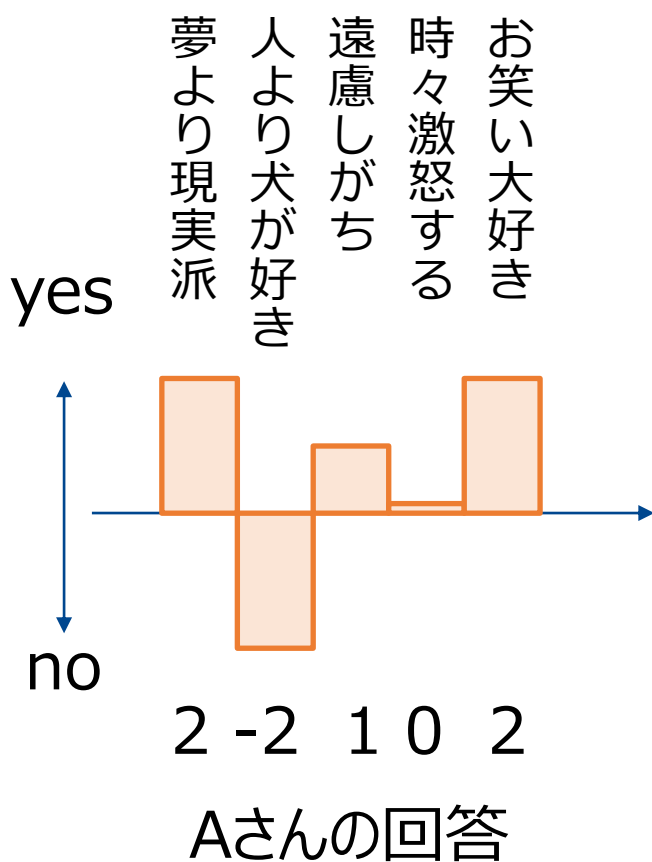
分解と合成, 一連の話をまとめると...(2/2)

(ふざけているように見えるかもしれませんが, これが本質!)

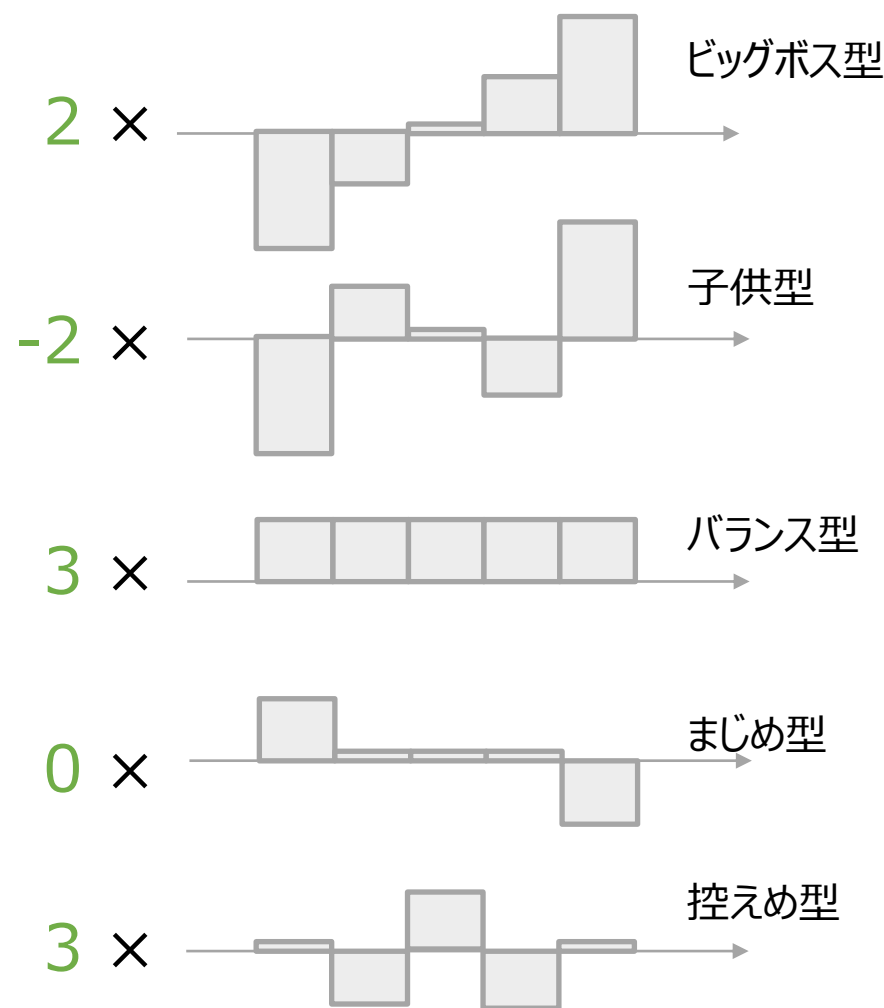
- 言い換えると, 成分量 $\alpha_1, \dots, \alpha_i, \dots, \alpha_d$ は, 材料セット $e_1, \dots, e_i, \dots, e_d$ から x を作るための「レシピ」



ベクトルの分解のありがたさをカレー以外の例で… ベクトル＝「性格診断5問の回答」場合



≡



※もちろんイイカゲンな例です

ここまでのおさらい

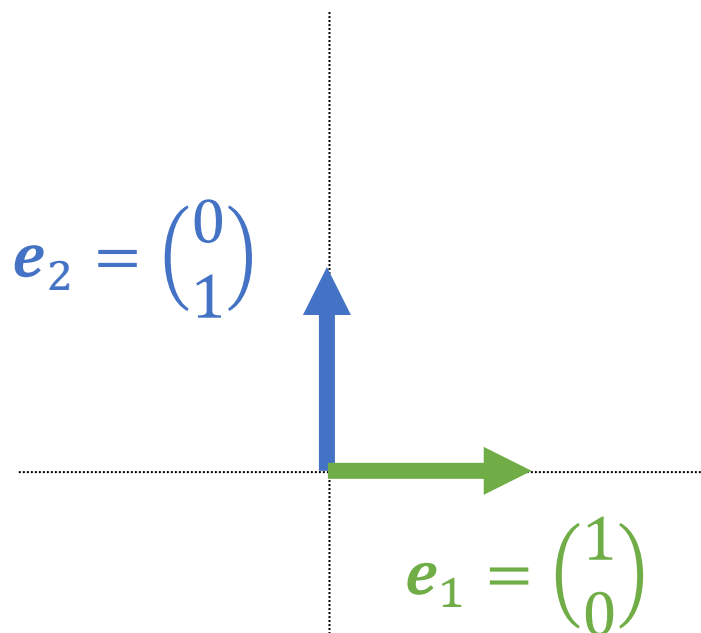
- 分析とは、いろいろな成分に**分解**することである
 - 分解することで、よくわかることがある
- 「ベクトルの分析」も同じ
 - ベクトル \mathbf{x} を「単位ベクトル群」に分解することで、理解する
- ベクトルの分析には「単位ベクトル」との**内積**が使える
- さらに**合成**すれば元通りに
 - ※ただし完全に元通りになるためには条件がある
- 分析の方法は一意ではない
 - 回転させた単位ベクトル群でも可能



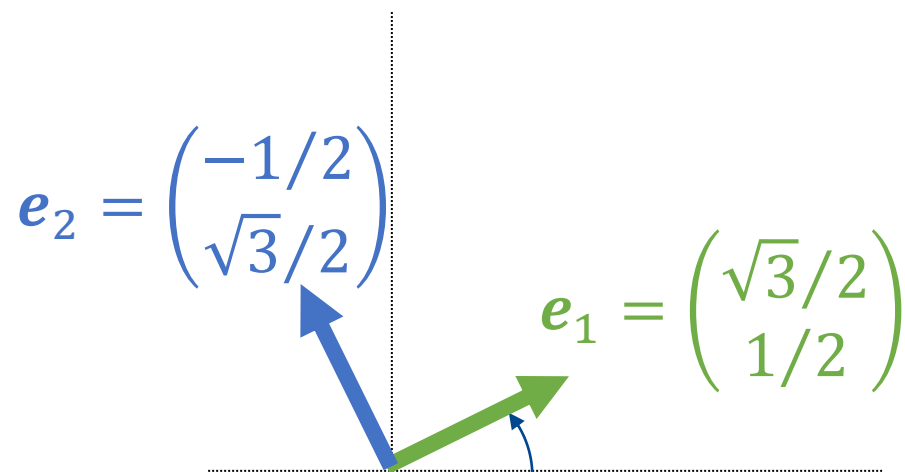
基底

あまり面白みもない「単位ベクトルのセット」,
でも実は「モノの見方」決めるのがこの基底

(数学的には厳密ではないが)
単位ベクトルのセットのことを「基底」と呼ぶ

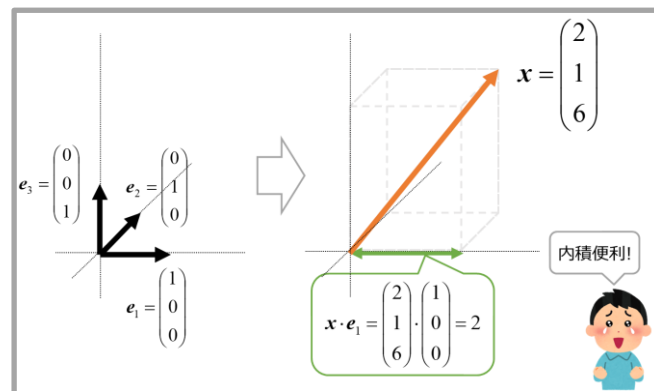
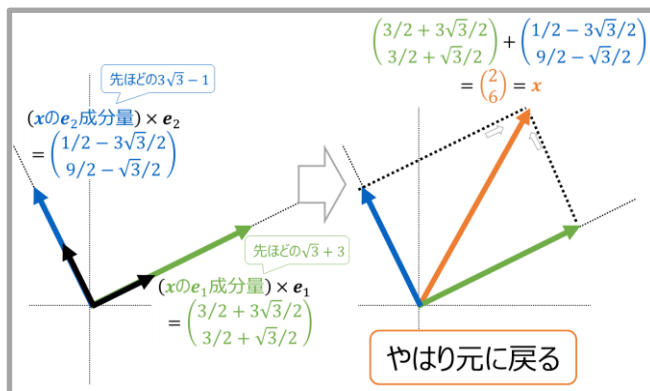
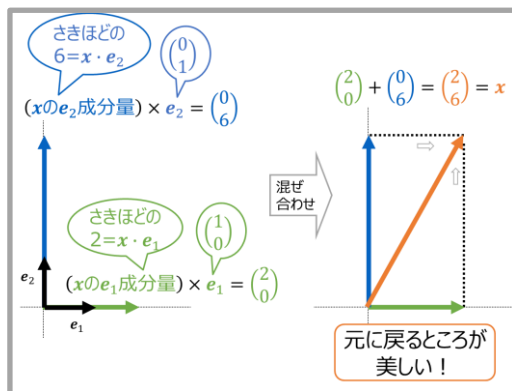


これはわかりやすい「基底」



こちらも「基底」

前述の通り、任意のベクトル x を「分解して・合成して・元に戻る」能力がある基底(単位ベクトルのセット)がある



共通する性質は？

1. d 次元ベクトル用には d 本のベクトル
2. 直交している(90度で交わる)
3. それぞれ長さが1

単位ベクトルのセットがこれら3つの性質を満たせば、「元に戻せる」基底になっている!

便利!



この便利な基底を「完備正規直交基底」と呼ぶ

早く言葉みたいですが...

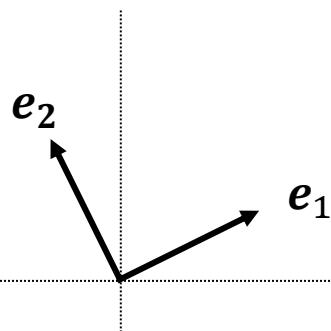


完備

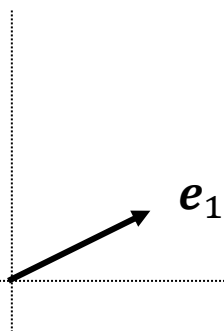
直交

正規

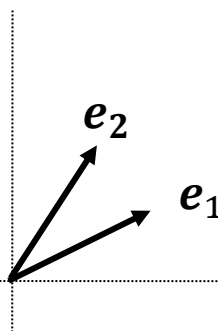
1. d 次元ベクトル用には d 本のベクトル
2. 直交している(90度で交わる)
3. それぞれ長さが1



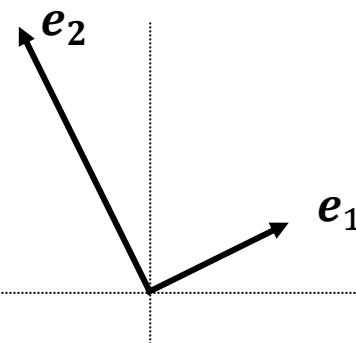
完備正規直交基底



完備でない



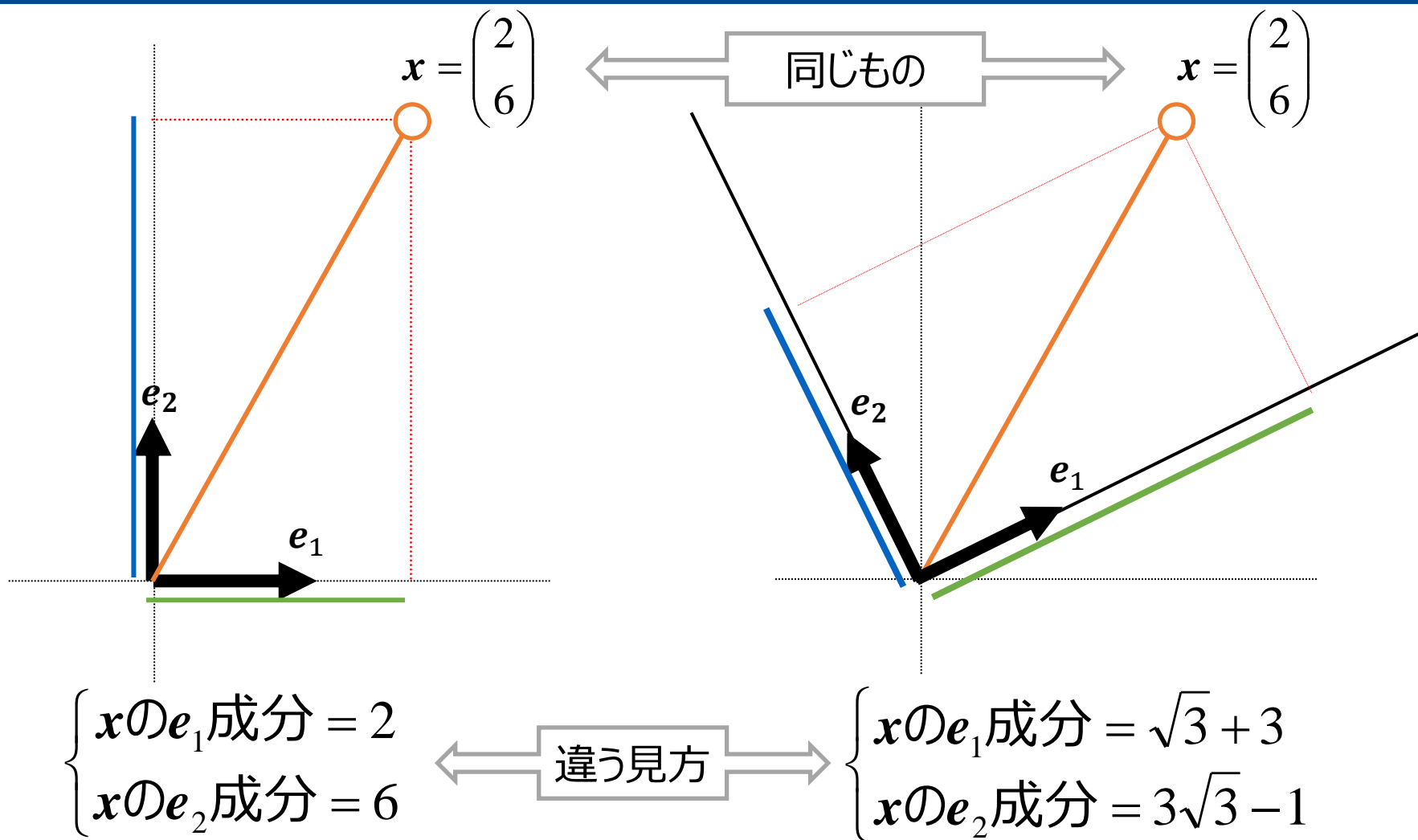
直交でない



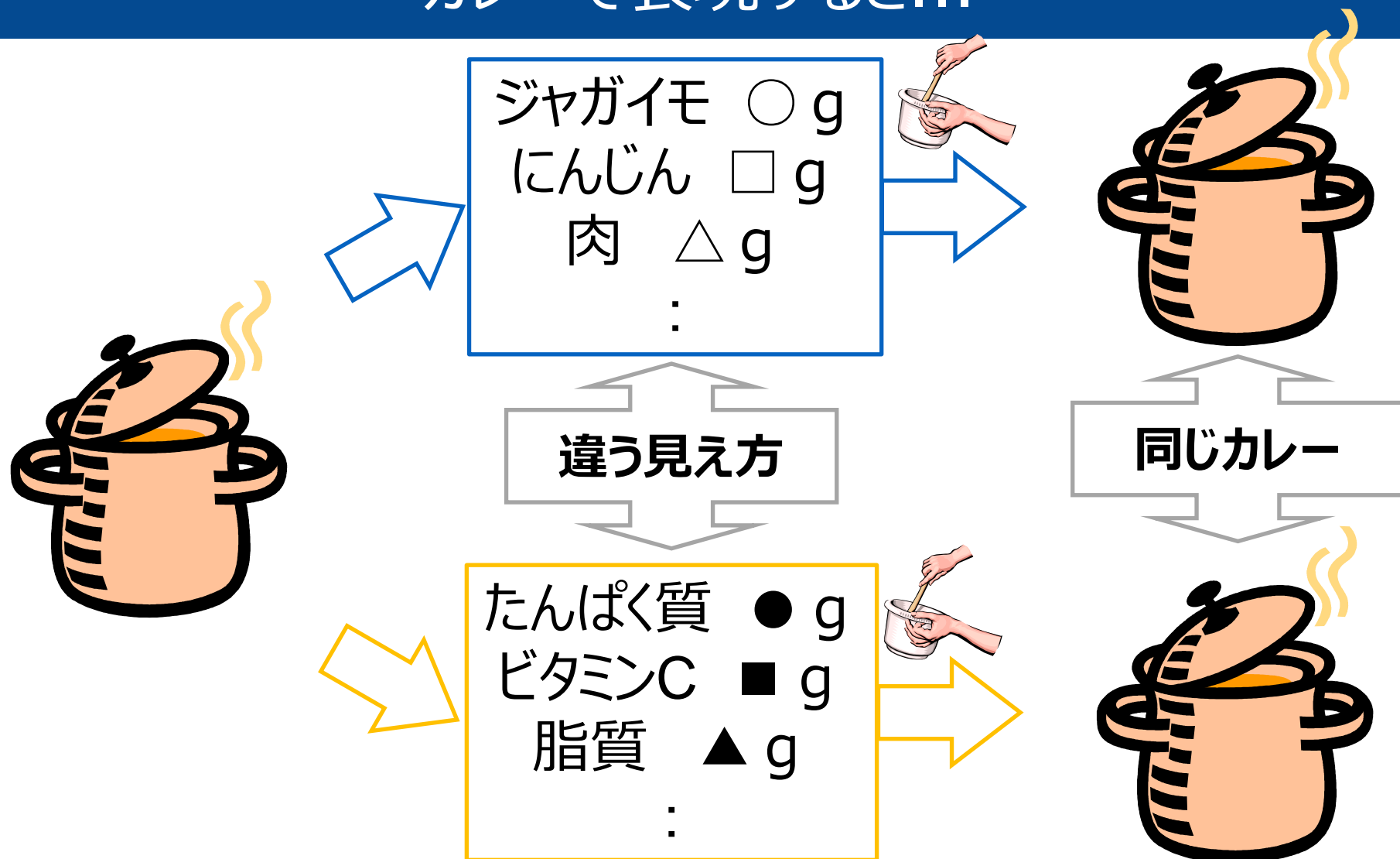
正規でない

詳細は付録!

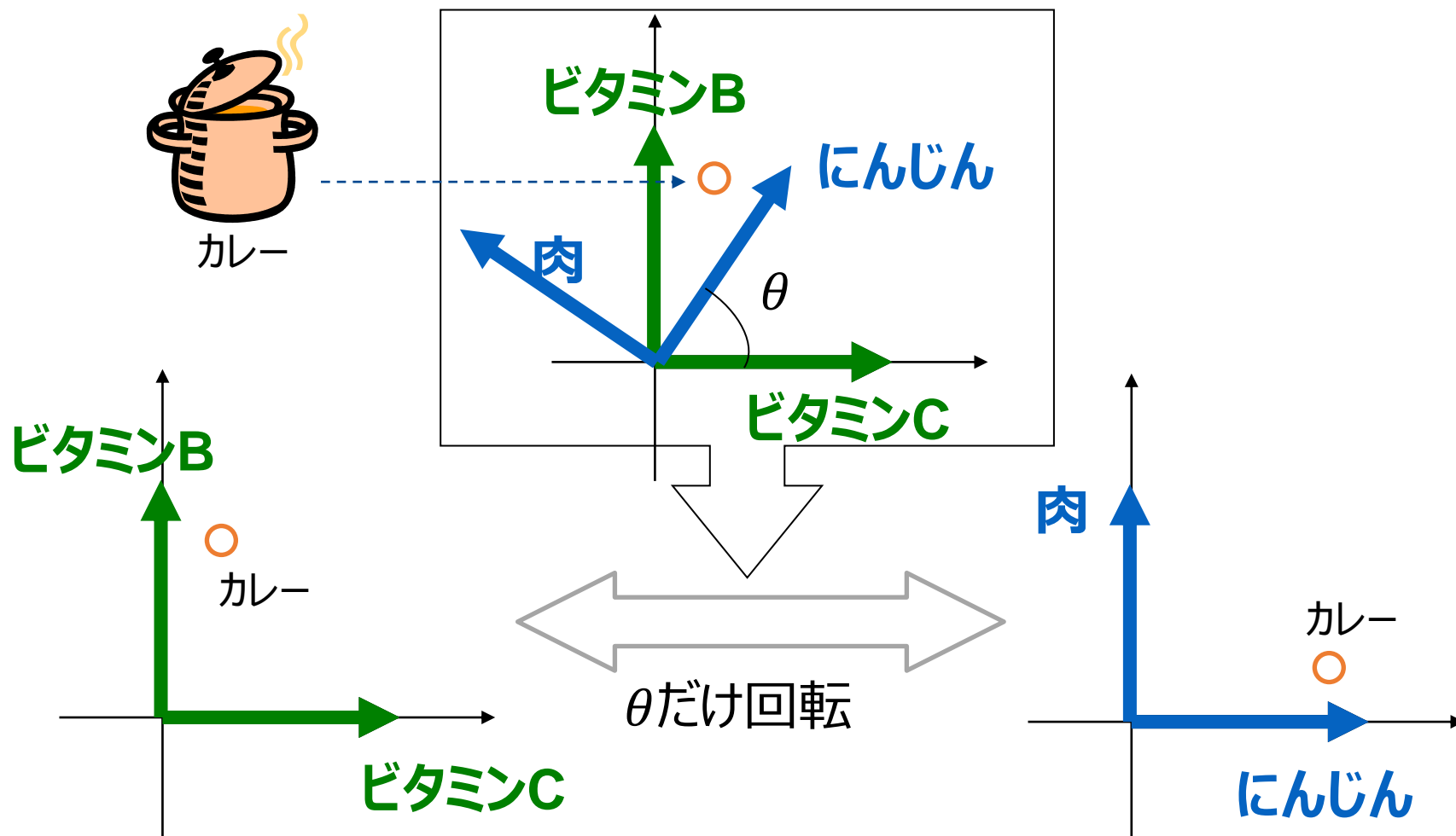
基底が変われば成分量が変わる： = ものの見方が変わる！



基底が変わればモノの見方が変わることを カレーで表現すると...



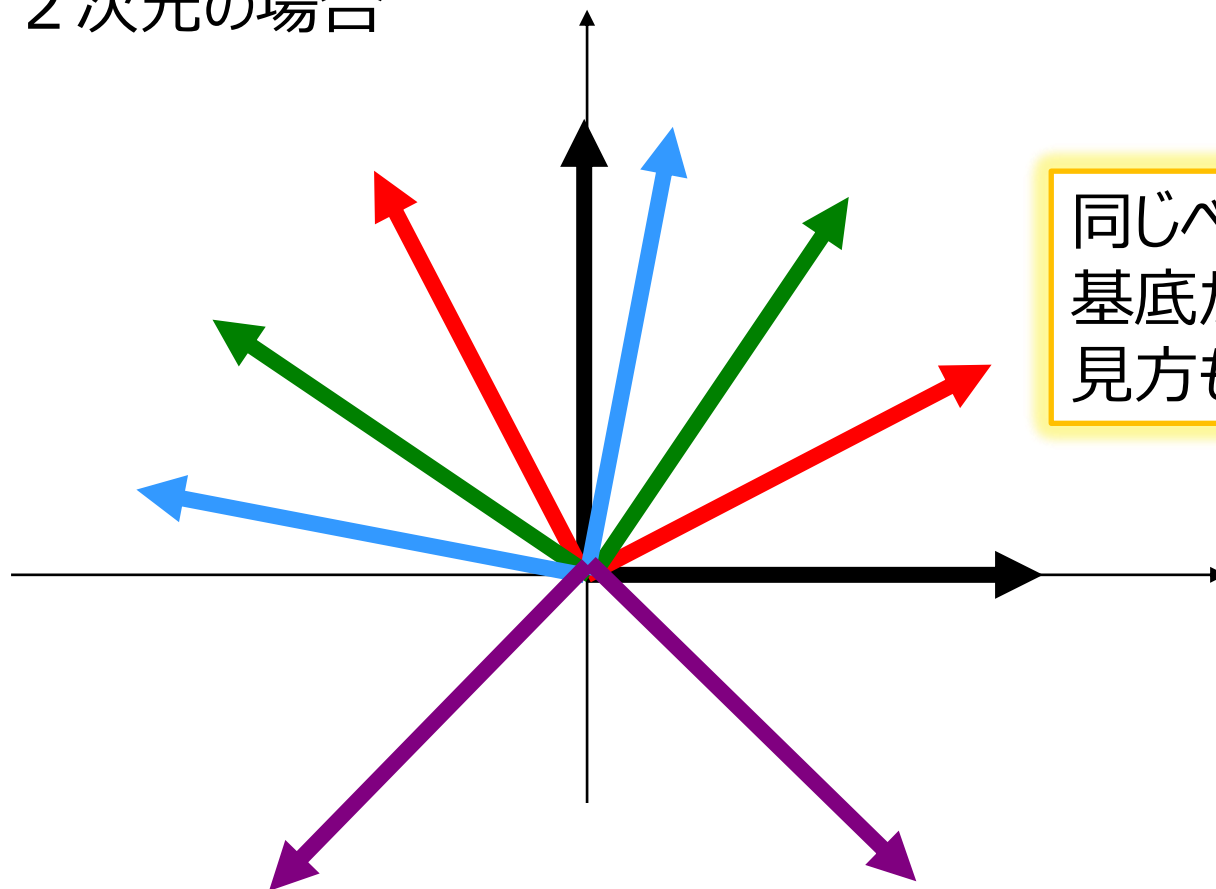
ある基底での分析結果を回転するだけで、
他の基底での分析結果が求まる



(2つの分析結果は1対1対応！)

回転により「完備正規直交基底」が無限にできる。
ということは、モノの見方(=分析の方法)も無限通り！

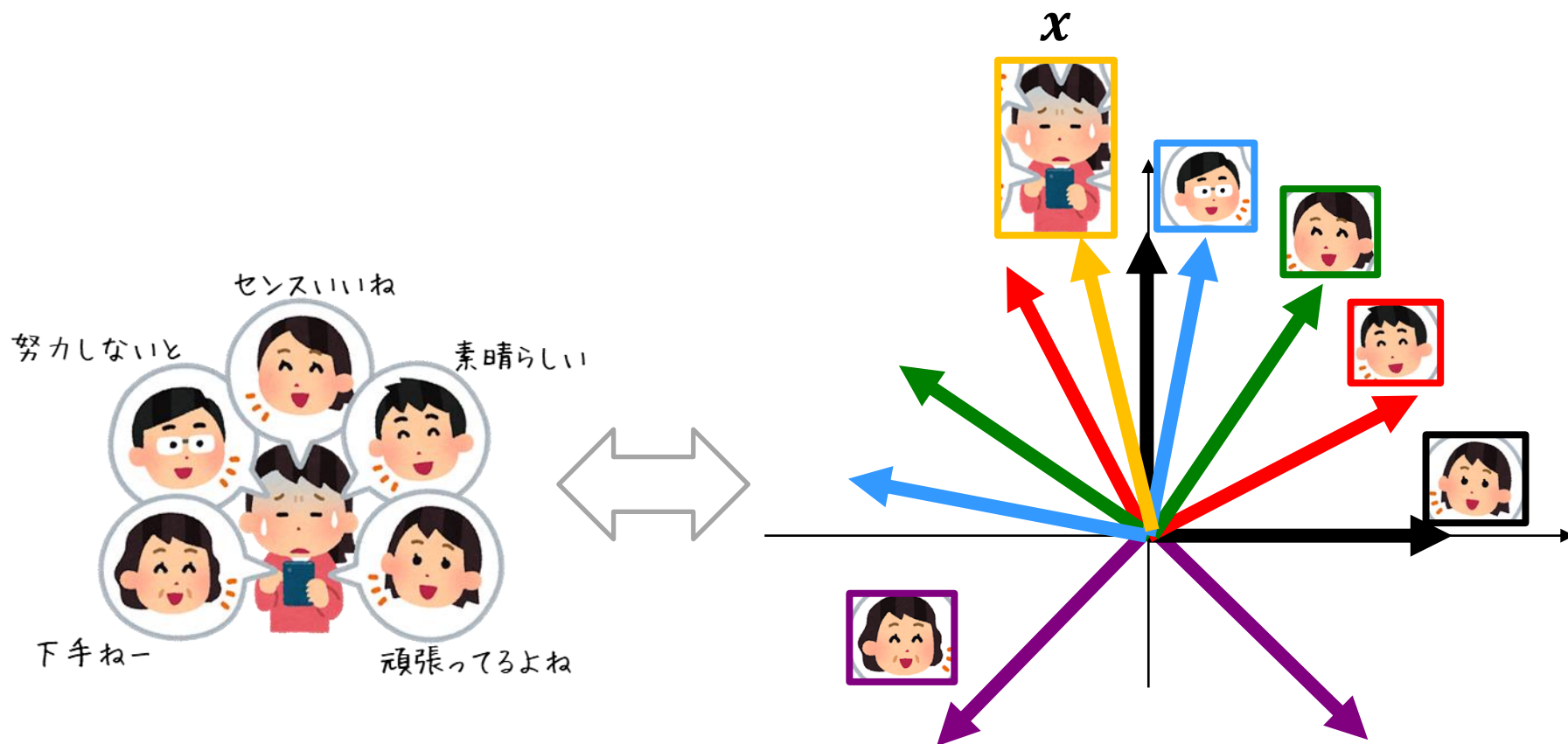
- 例：2次元の場合



同じベクトルでも
基底が変われば
見方も変わる！

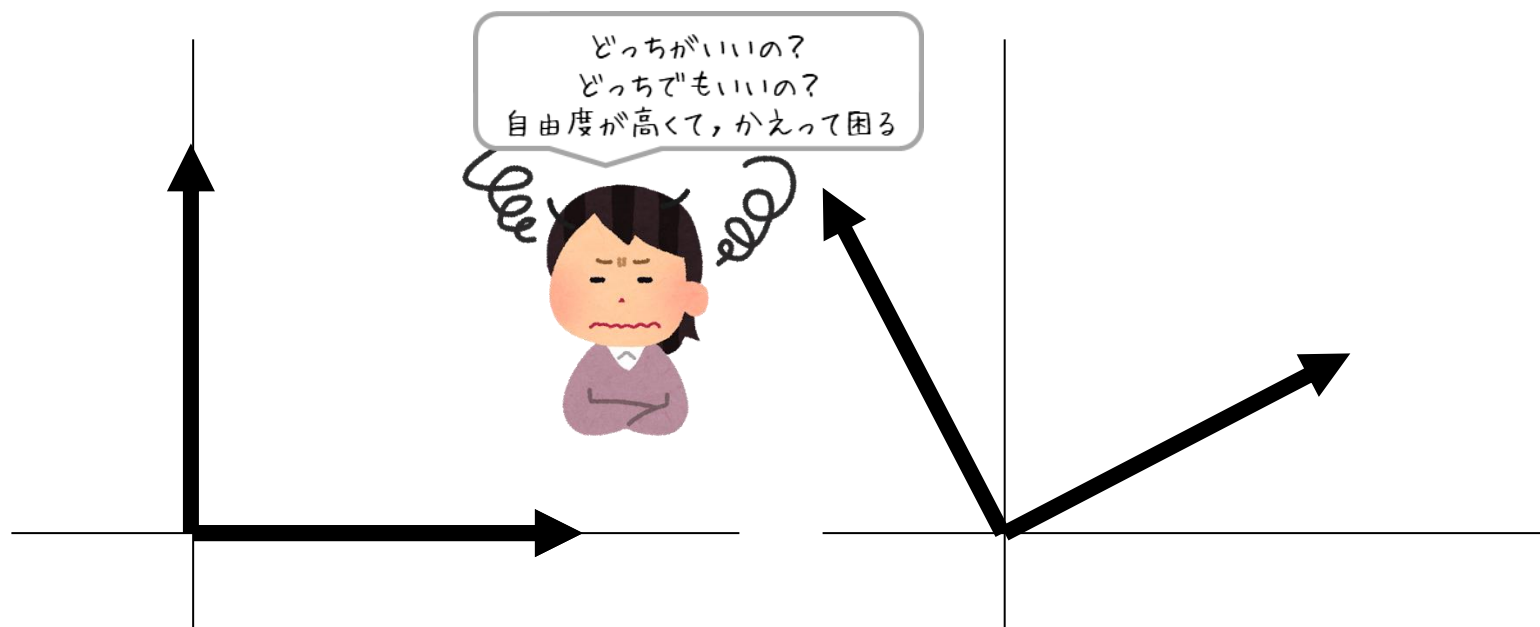
似たようなことは日常でもある！

人によって、同じものが違って見える



逆に困った！ 各分析課題にとって「よい」基底とは何かを考えないと...

- 「どれがよい？」と言われても、どの正規直交基底も任意のベクトルを(分解・合成して)元に戻せる点では同じ...



- でも、「いま分析のターゲットとしている」ベクトル集合(=データ集合)にとって「よい基底」なら決められるかも...

コンパクトな分析結果を与える「基底」



分解して合成しても、完全には元に戻らない。
でも、そのおかげで、よりシンプルな分析結果が求まる

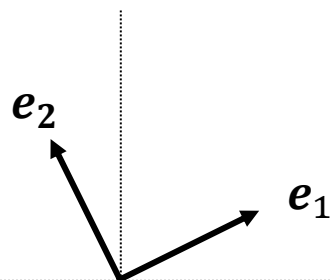
「完備正規直交基底」をもう一度

完備

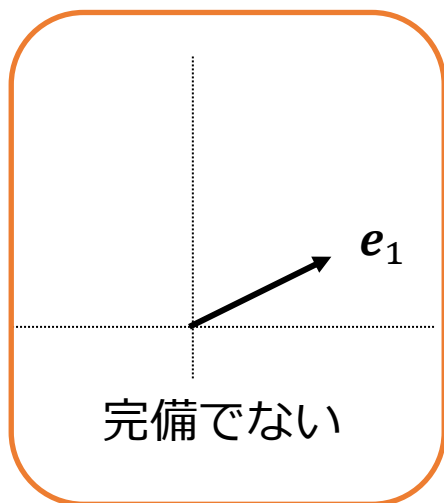
直交

正規

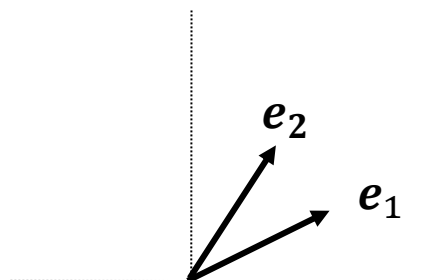
1. d 次元ベクトル用には d 本のベクトル
2. 直交している(90度で交わる)
3. それぞれ長さが1



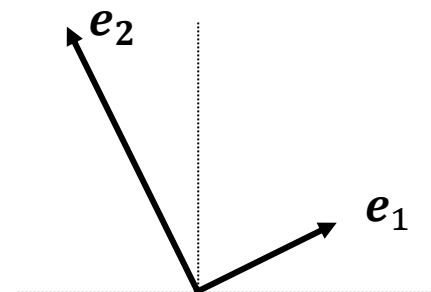
完備正規直交基底



完備でない



直交でない

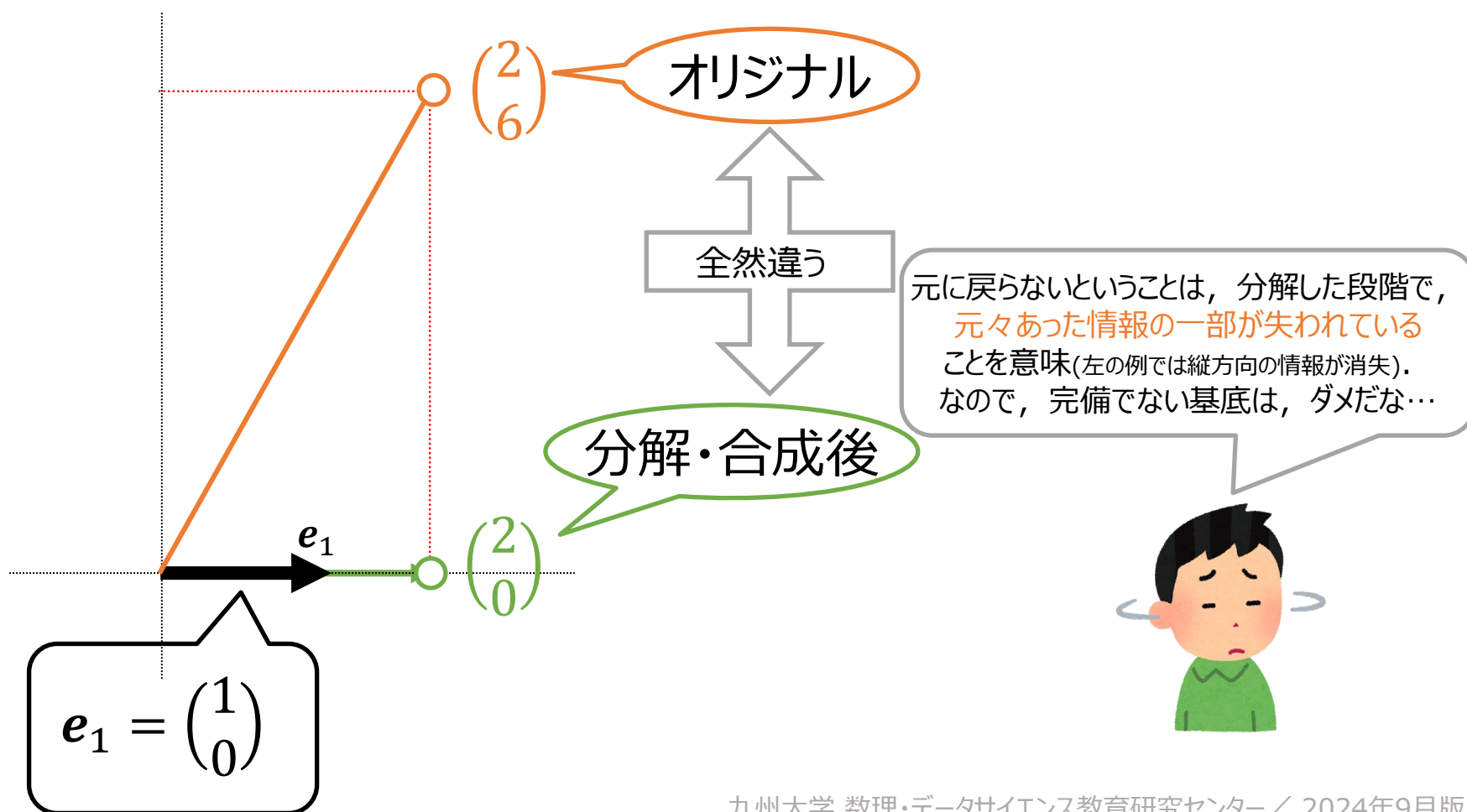


正規でない

ダメな基底なのか？

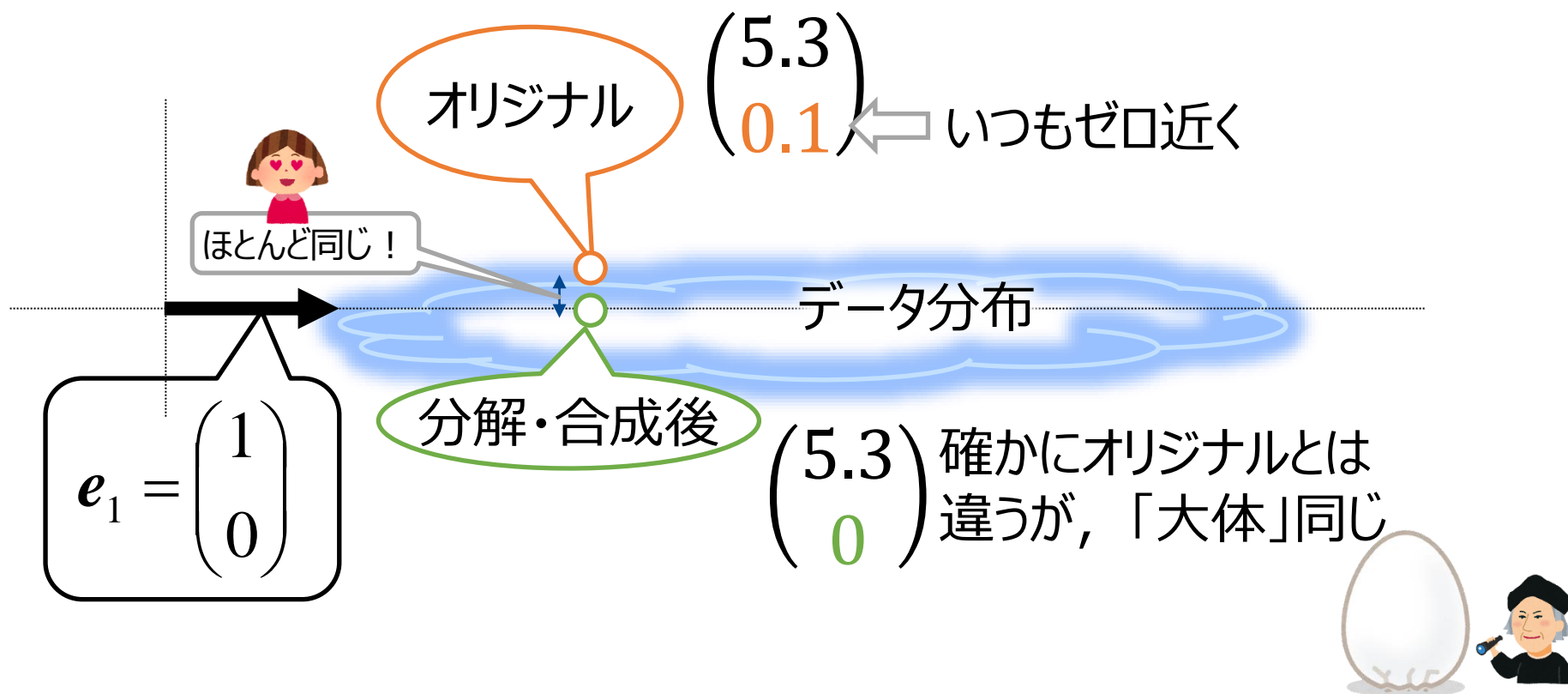
完備でない基底を使うとどうなるのか？ → もとに戻らない

- ex. 2次元で基底が1つしかないと，もとに戻らない！



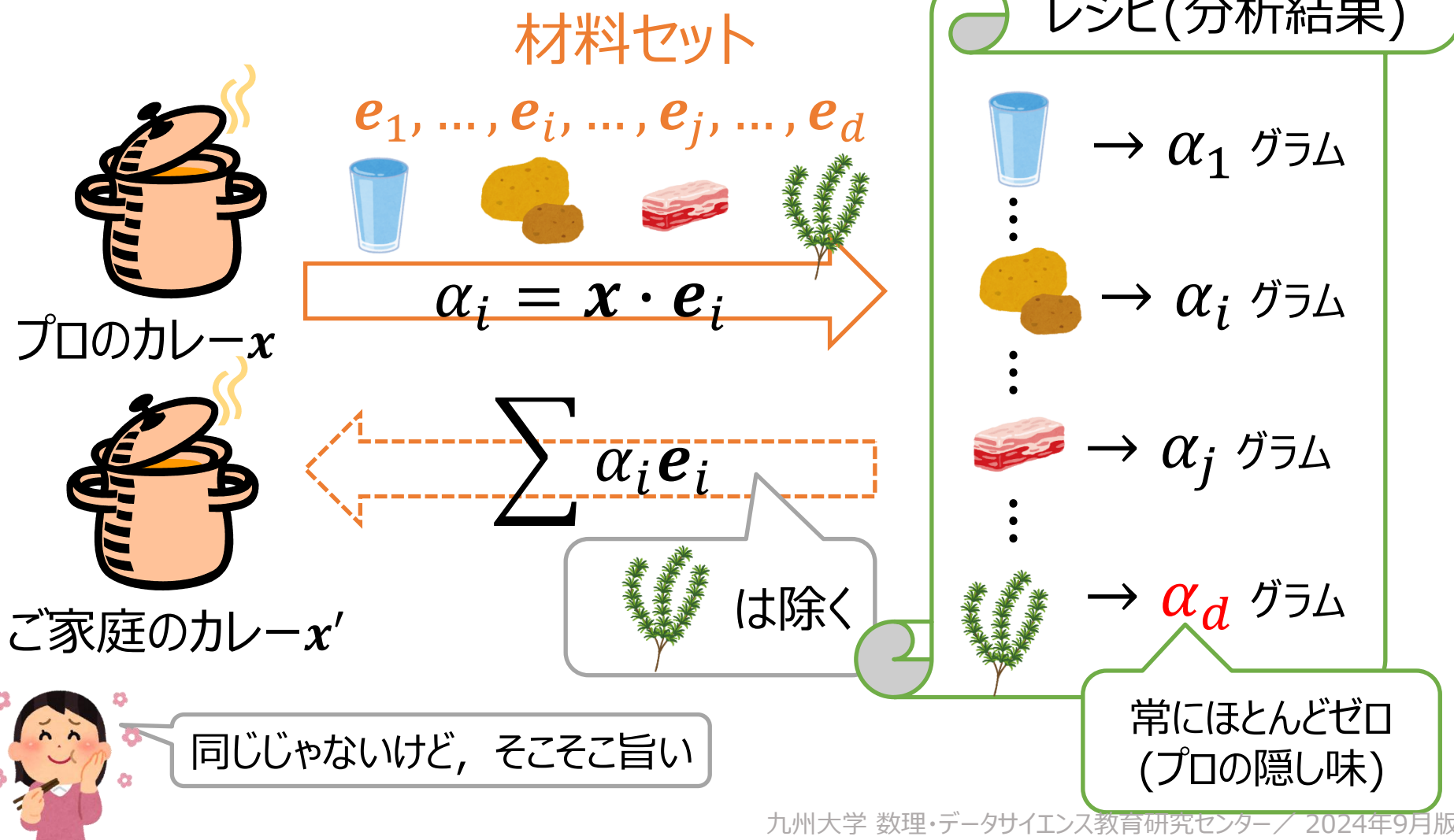
しかし！ データによっては全部の単位ベクトルを使わなくても「ほぼ」OKなケースもある

- もしデータが第1軸付近にしかないのなら，1個の単位ベクトルだけで「ほぼ」OK？



注：コロンブスの卵

カレーの例で説明すると(1/2) : プロの味でなくても, ご家庭の味でOKなケース!



カレーの例で説明すると(2/2)： 「どこで手を抜くか？」もセンスの一つ




材料セット

$e_1, \dots, e_i, \dots, e_j, \dots, e_d$



$$\alpha_i = x \cdot e_i$$

$$\sum \alpha_i e_i$$

大事なを除いてしまうと



絶対に必要なやつ
(いつも量が多い)



→ α_1 グラム

⋮



→ α_i グラム

⋮



→ α_j グラム

⋮



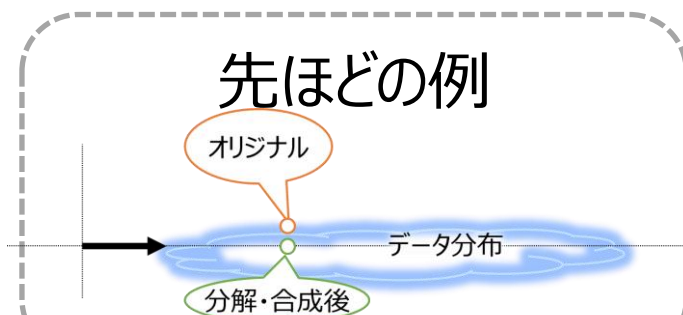
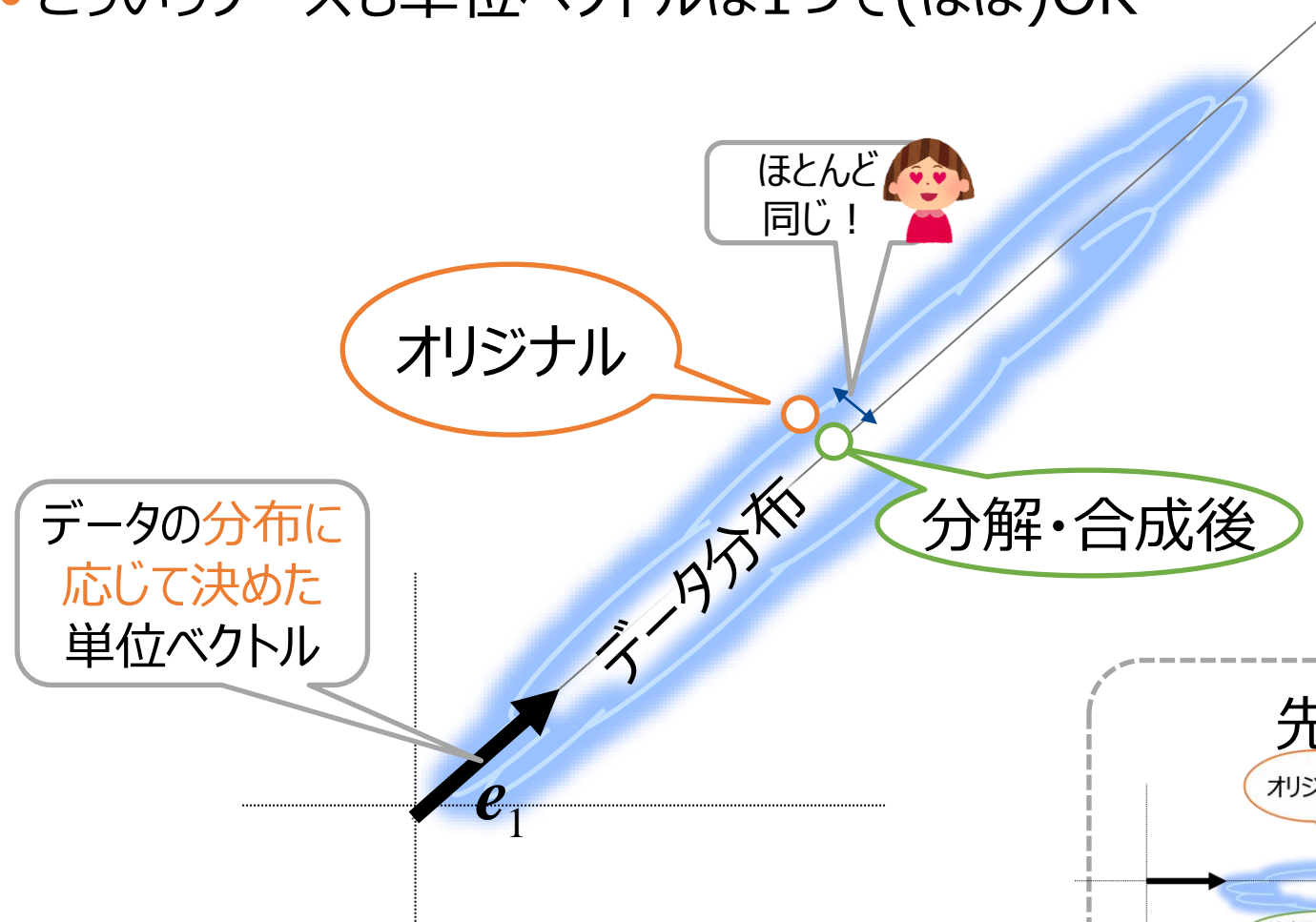
→ α_d グラム



全然違う～！こりゃだめだ...

全部の単位ベクトルを使わなくてももソコソコ戻る, より一般的なケース

- こういうケースも単位ベクトルは1つで(ほぼ)OK



「そこそこでOK」とするメリット

- d 次元ベクトルを, d 個未満の成分でコンパクトに表現

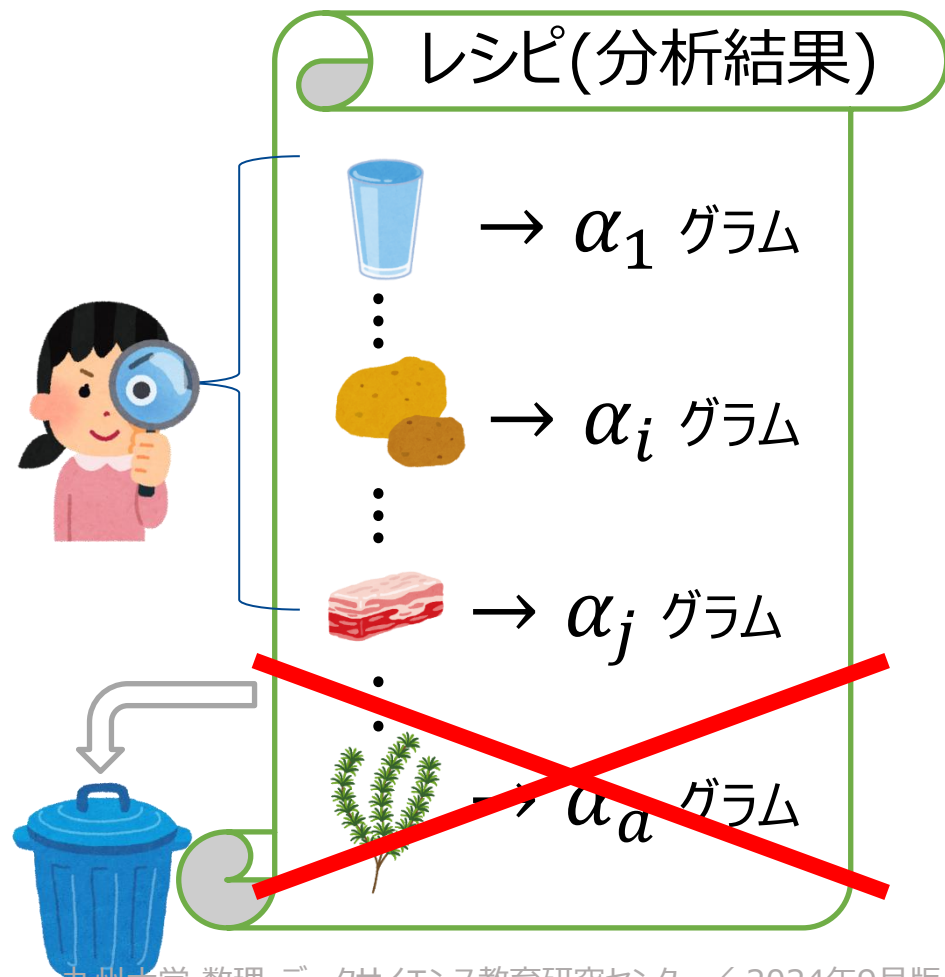
- したがって :

- 主要な成分だけを概観できる !

- = 主成分分析のアイデア

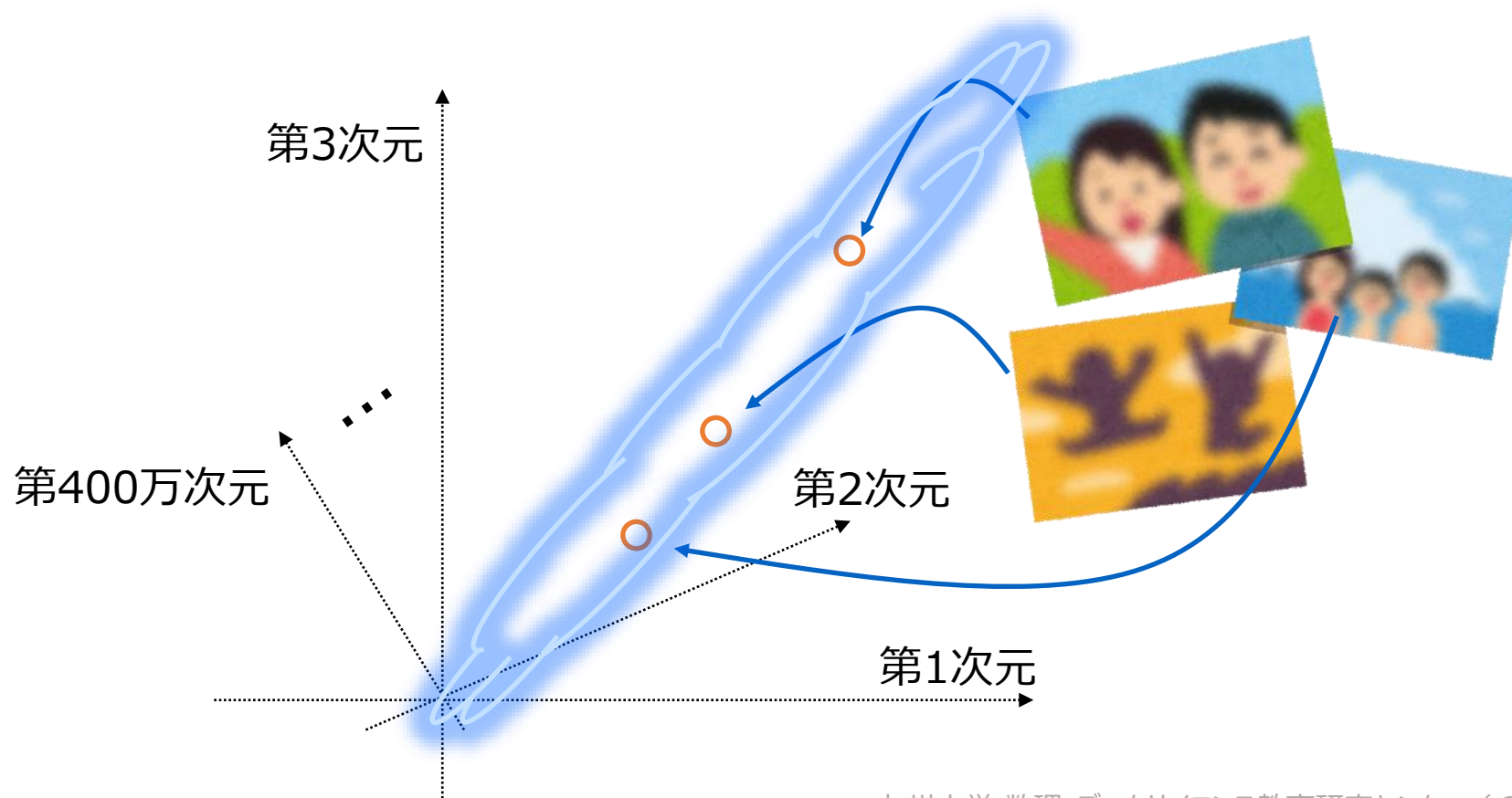
- 主要でない成分を捨てることで,
データ圧縮にもなる

- d 次元ベクトルの情報を送る場合,
 d 個の数字をそのまま送らずに,
より少ない数で送ることができる



画像が圧縮できるのは、実はこの性質を使っています

- 地デジやデジカメ写真はこの原理で情報圧縮(=コンパクト化)
- 画像は実は偏って分布している！



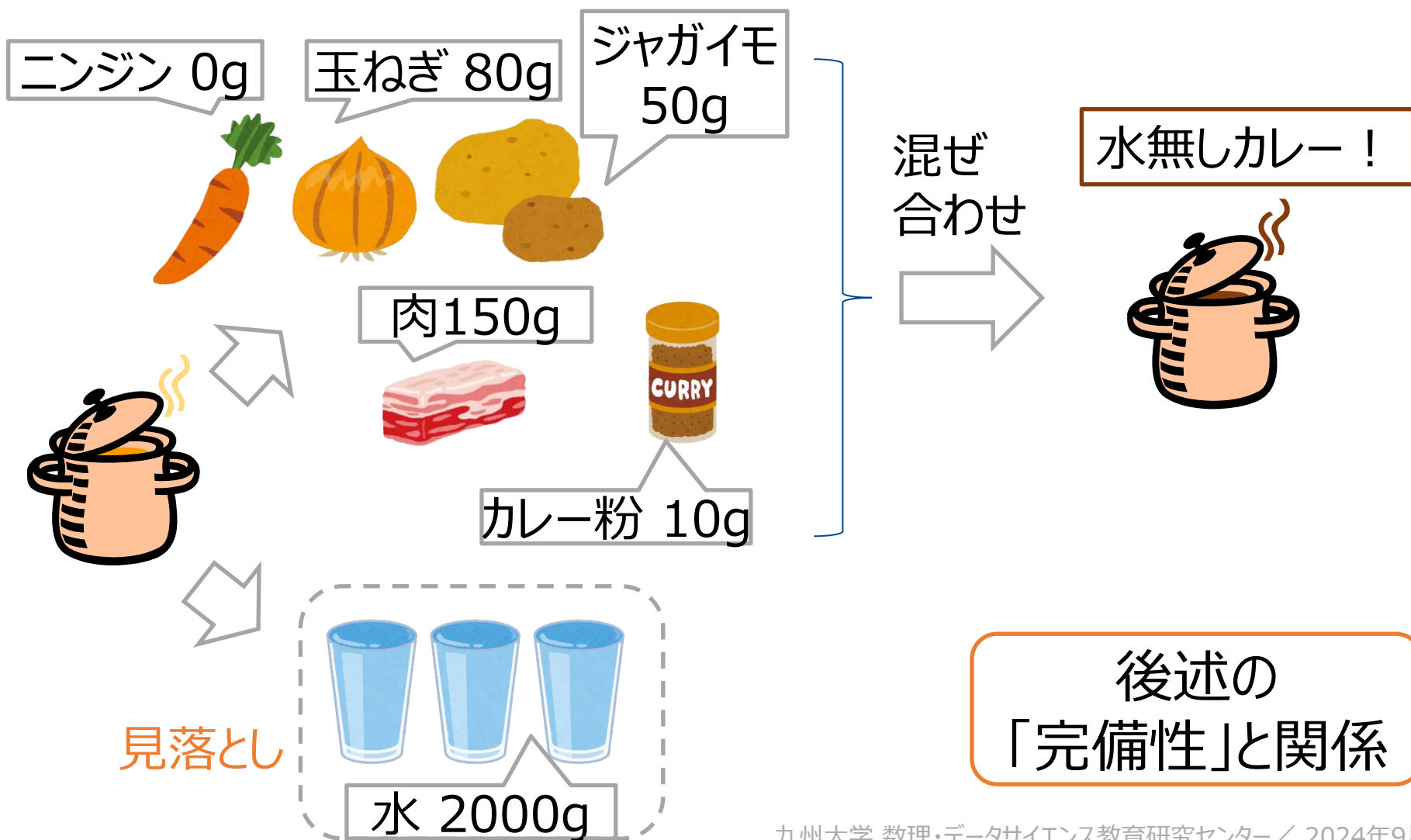
【付録 1】 どのような基底を用いるべきか？

「分解して合成すれば元に戻る」ための条件



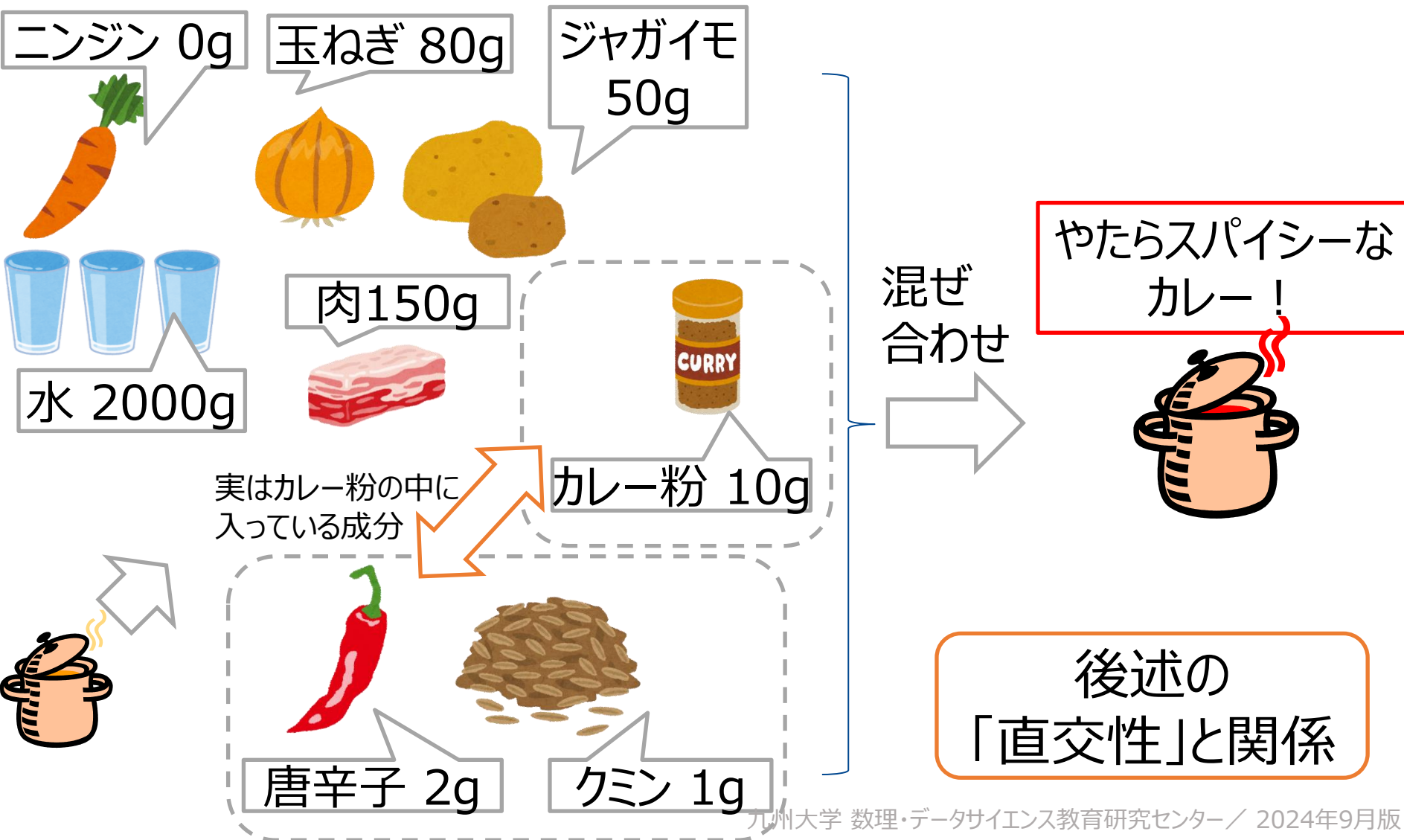
分析の際にケアすべきポイント(1/4)

基本的に見落としはNG



分析の際にケアすべきポイント(2/4)

分析項目に重複がないほうが良いだろう



分析の際にケアすべきポイント(3/4)

分析する単位は統一したほうがよいだろう

ニンジン 0本

玉ねぎ 80g

ジャガイモ
30 cm^3

水 3カップ

肉 0.15 kg

比べにくい！



カレー粉 10000 mg

元のカレーには
完全に戻るのだが...

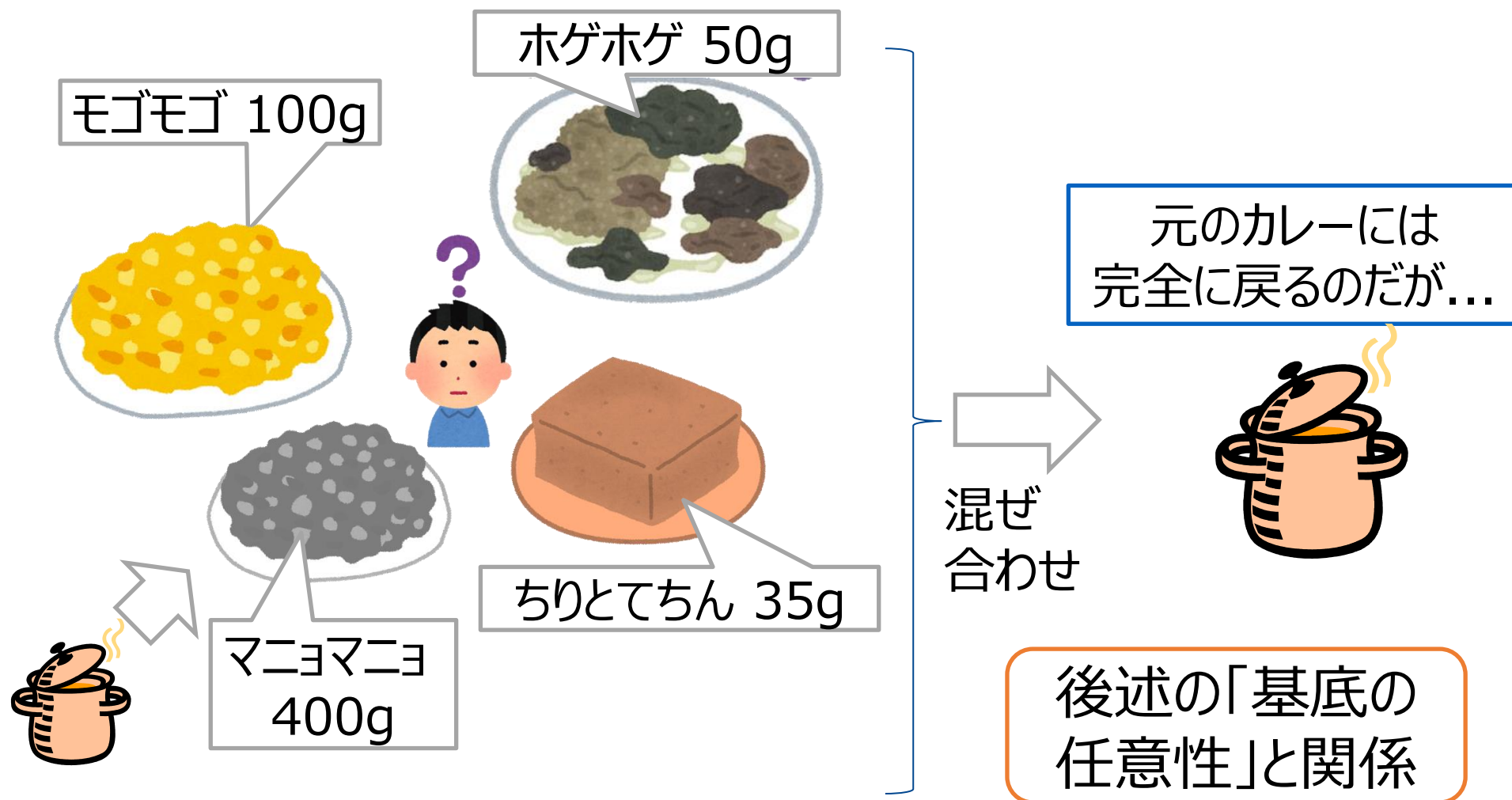
混ぜ
合わせ



後述の
「正規性」と関係

分析の際にケアすべきポイント(4/4)

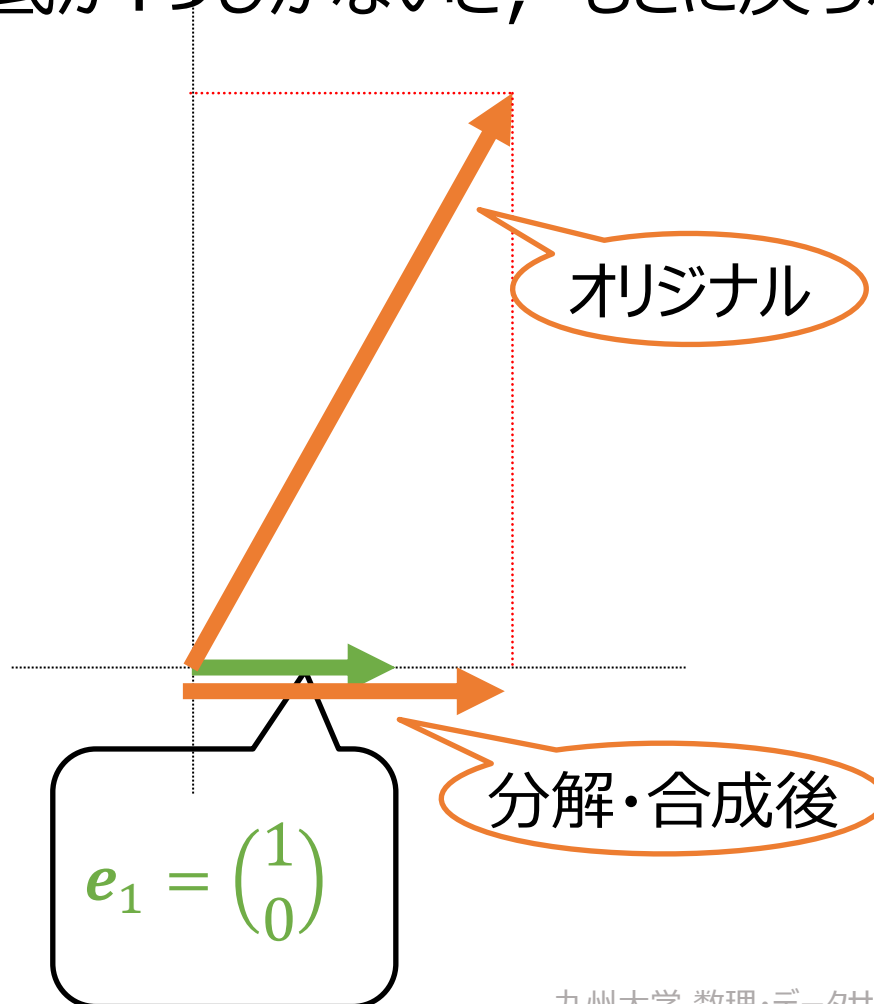
解釈容易な成分に分解したほうがよい



「分解して合成すれば元に戻る」ための条件①

基底の完備性(1/2)

- 2次元で基底が1つしかないと、もとに戻らない！



水無しかレー！

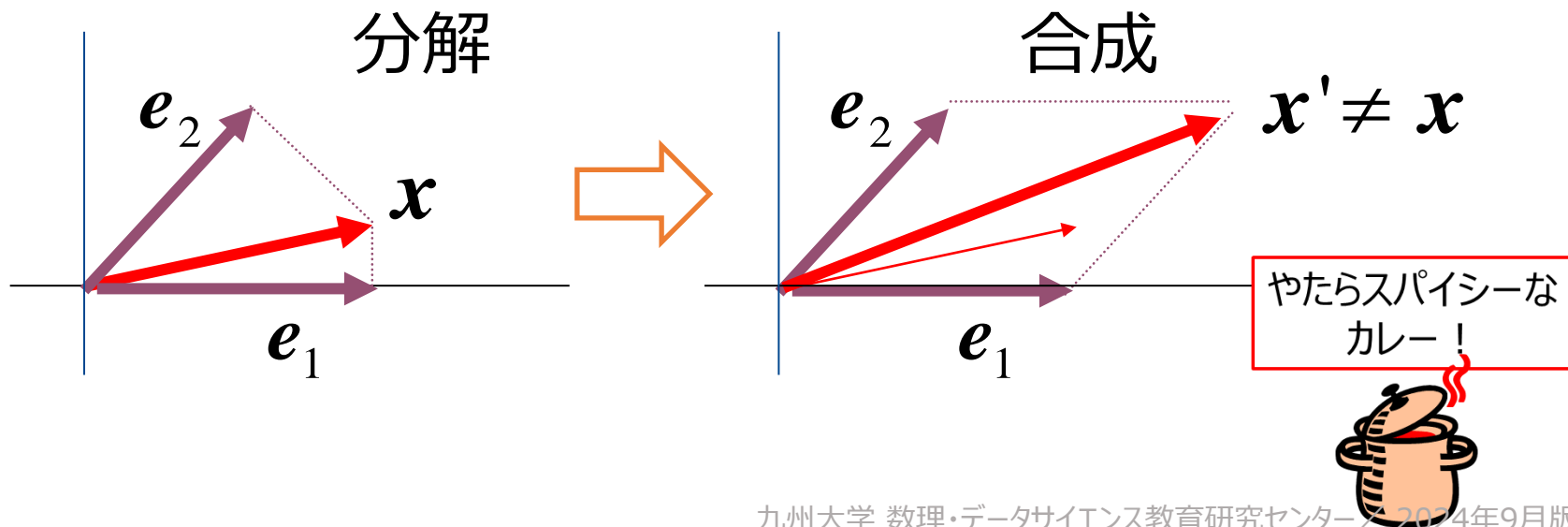


「分解して合成すれば元に戻る」ための条件① 基底の完備性(2/2)

- 任意の2次元ベクトル x は, 2個の2次元単位ベクトルによる分解→合成の結果, 元の x に戻る
- 任意の3次元ベクトル x は, 3個の3次元単位ベクトルによる分解→合成の結果, 元の x に戻る
- 任意の d 次元ベクトル x は, d 個の d 次元単位ベクトルによる分解→合成の結果, 元の x に戻る
- この状況を「完備」と言います
(ほんの少し数学的にはイイカゲンなことを言っていますが, 気にしない, 気にしない)

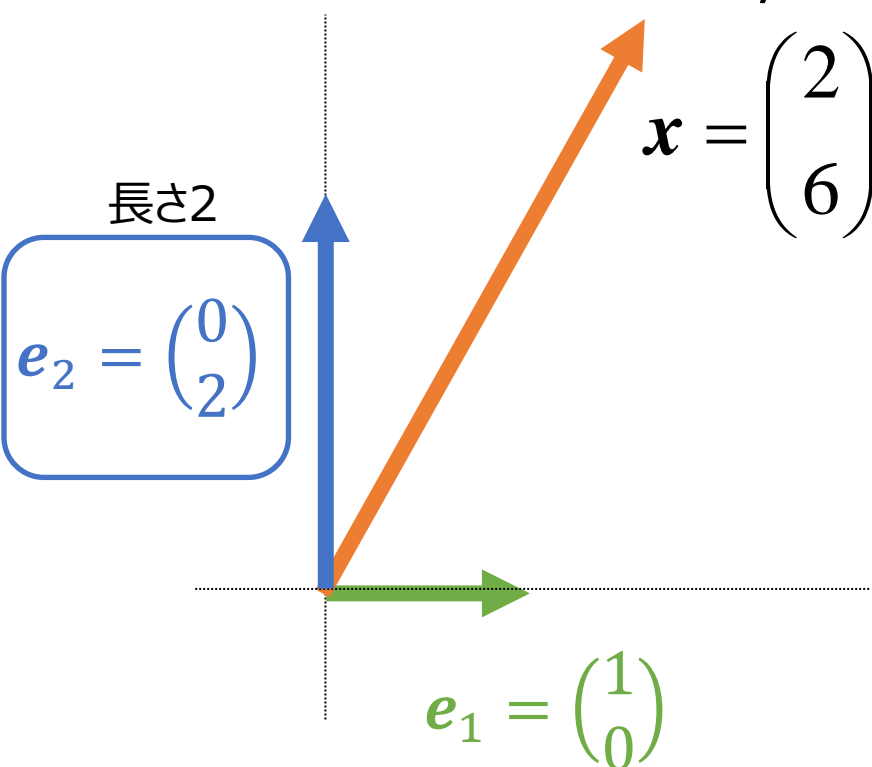
「分解して合成すれば元に戻る」ための条件② 基底の直交性

- 単位ベクトルが互いに90度で交わっていること
 - 高校の時に習ったかもしれませんが, $e_i \cdot e_j = 0$ (ただし $i \neq j$)
- 非直交だと, 分解合成で元に戻らない!
 - 頑張れば戻せないこともないが, かなり面倒くさい



「分解して合成すれば元に戻る」ための条件③ 基底の正規性

- 単位ベクトルの長さは、常に 1 で！



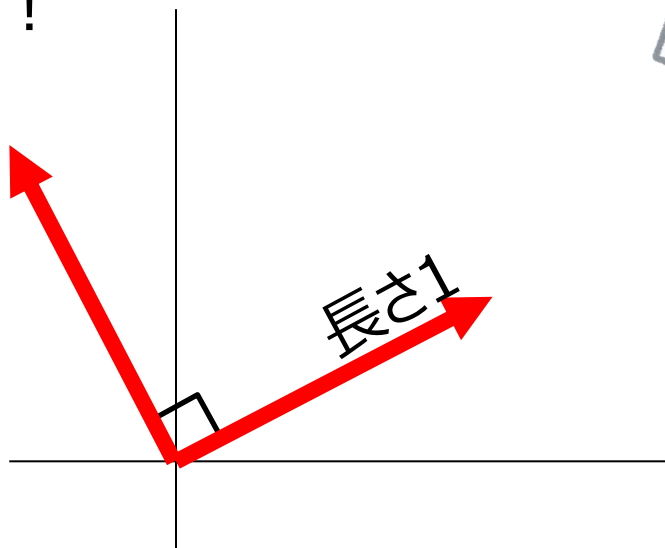
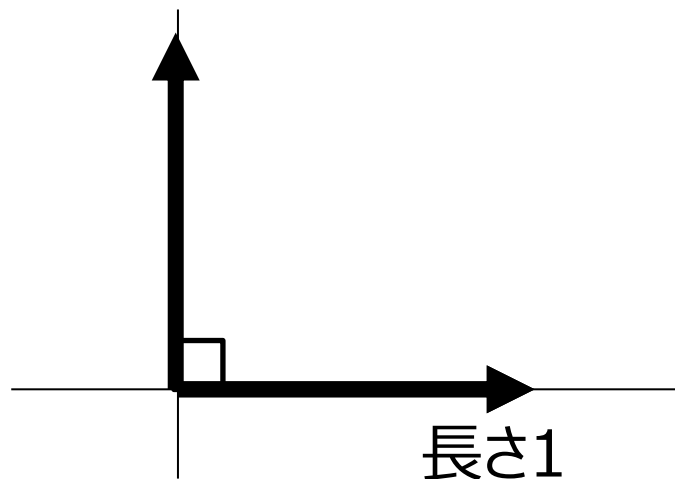
$$\begin{cases} x \text{ の } e_1 \text{ 成分} = x \cdot e_1 = 2 \\ x \text{ の } e_2 \text{ 成分} = x \cdot e_2 = 12 \end{cases}$$

???

e_2 が 12 も入ったら第 2 軸は $12 \cdot \begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 24 \end{pmatrix}$

以上 3 つの条件を満たす単位ベクトルの組を 「完備正規直交基底」と呼ぶ

- 以下はいずれも2次元の完備正規直交基底
 - =どちらもカレーは元通り！



- もちろん3次元以上でも同様

