Data Driven Process Engineering

Sampling-Based Data Driven Optimization

Fani Boukouvala, Ph.D.

Assistant Professor

School of Chemical and Biomolecular Engineering

Positive Spanning Sets

Learning Objectives

- Define the importance of geometry of samples
- Define span and positive spanning sets
- Explain difference between positive spanning set and basis
- Illustrate a positive spanning set/basis in 2D
- Explain positive spanning set properties

Sampling

- All data-driven optimization algorithms start with **samples (data)**

- When sampling locations can be “designed”, many rules exist linked to theory and convergence on where they should be located

- In local direct-search methods, this all starts with positive spanning sets and bases

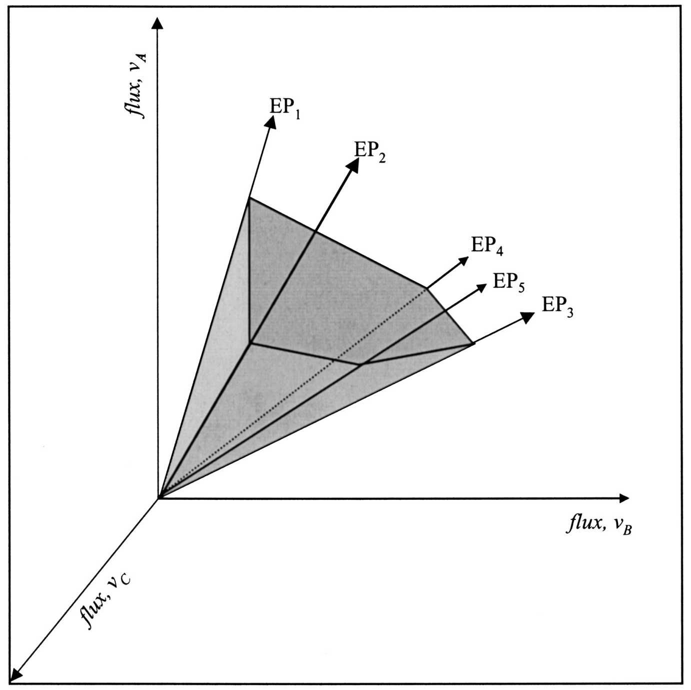
Positive Spanning Sets/Bases

- Term “Span” or “Linear Span”: smallest linear subspace that contains a set of vectors

- Positive span of a set of vectors $[v_1, v_2, \ldots, v_r]$ in R^n

is the **convex cone**:

$$\{v \in R^n : v = a_1 v_1 + a_2 v_2 + \cdots + a_r v_r, \qquad a_i \geq 0 \; for \; i = 1, \ldots, r\}$$

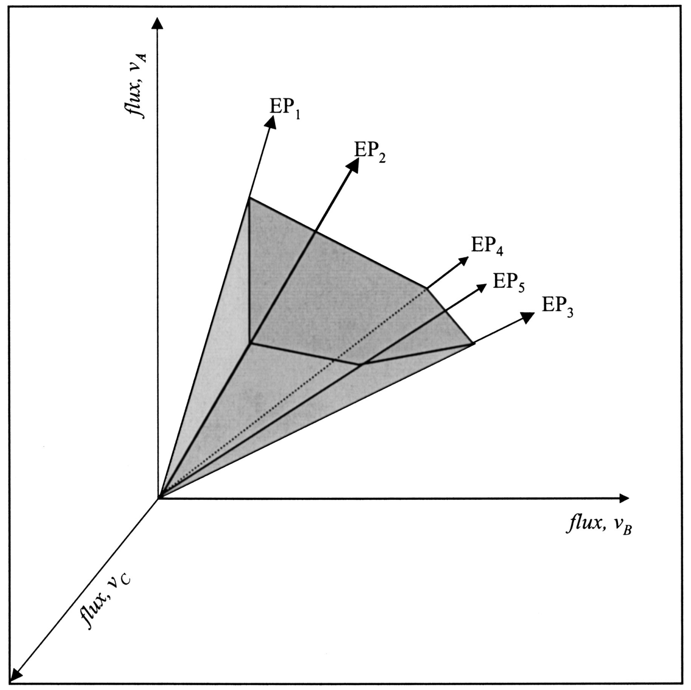


Positive Spanning Bases

- Positive span of a set of vectors $[v_1, v_2, \ldots, v_r]$ in R^n

is the **convex cone**:

$$\{v \in R^n : v = a_1 v_1 + a_2 v_2 + \cdots + a_r v_r, \qquad a_i \geq 0 \; for \; i = 1, .., r\}$$



- A **<u>positive spanning set</u>** in R^n is a set of vectors whose positive span is R^n

- A positive spanning set whose vectors are independent is a **<u>positive basis</u>**
  - **In a positive basis: no vector can be a linear (positive) combination of other vectors**

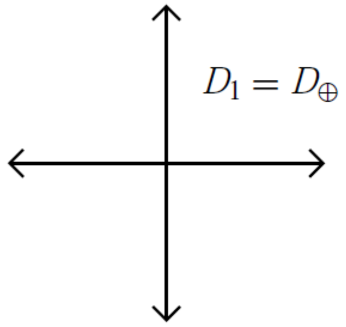
- A positive basis contains at least n+1 (minimal) and up to 2n (maximal) vectors

Positive Spanning Sets/Bases in 2D

- Contains at least n+1 (minimal) and up to 2n (maximal) vectors

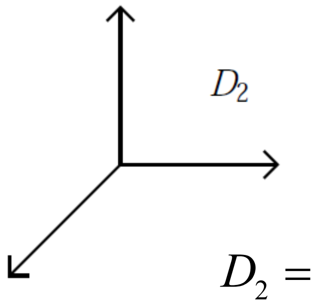
Example:
n=2

$$D_1 = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}$$



Maximal positive basis

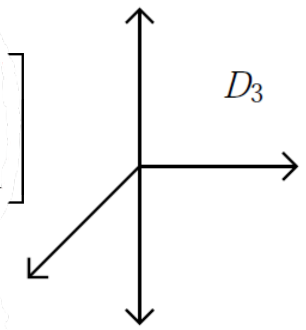
Minimal positive basis



$$D_2 = \begin{bmatrix} 1 & 0 & -\sqrt{2}/2 \\ 0 & 1 & -\sqrt{2}/2 \end{bmatrix}$$

$$\begin{bmatrix} 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \sqrt{2}\begin{bmatrix} -\sqrt{2}/2 \\ -\sqrt{2}/2 \end{bmatrix}$$

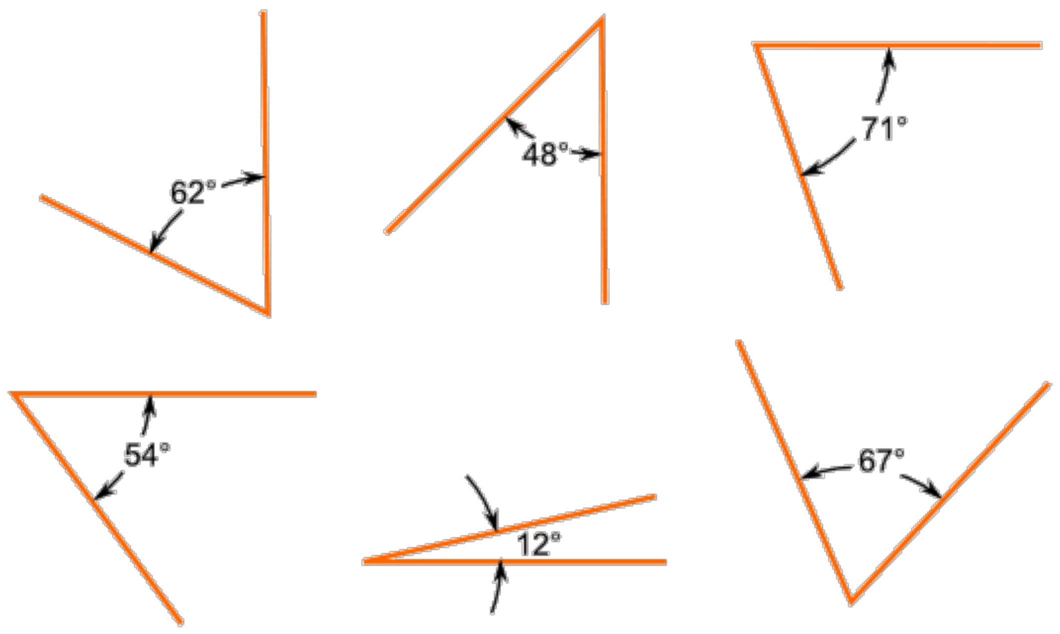
$$D_3 = \begin{bmatrix} 1 & 0 & -\sqrt{2}/2 & 0 \\ 0 & 1 & -\sqrt{2}/2 & -1 \end{bmatrix}$$



Positive spanning set

Importance of Geometry for Descent

- Property of positive spanning sets:
  - Given any non-zero vector v in R^n, there is at least one vector d in the positive spanning set such that v and d form an acute angle.
- Suppose that vector v is the negative gradient of a function f: $\left(-\nabla f(x)\right)$
- f is continuously differentiable
- Any vector d which forms an acute angle with $-\nabla f(x)$ *is a descent direction.*



- Descent direction for f at x means that:

There exists $a > 0$ such that:
$f(x + ad) < f(x)$

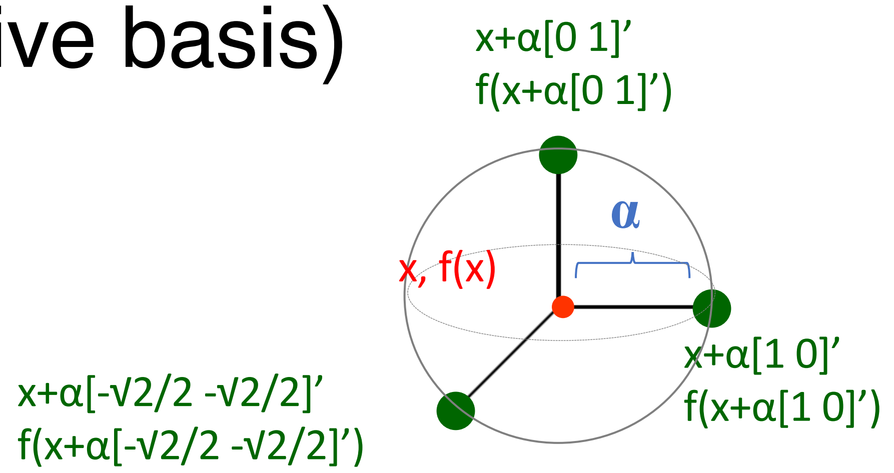
Importance of Sampling Design

- When building sampling-based optimization solver for expensive functions:
  - Guarantee improving obj function at each step, **BUT** not sample more than necessary
  - If the sampling region (sample vectors) form positive spanning basis, then we can guarantee that descent direction is found.
  - No need than more than **n+1** vectors at each step.

- Assume we sample around a point x on *n+1* points around x in the form

$$x + \alpha d \text{ where } \alpha > 0$$

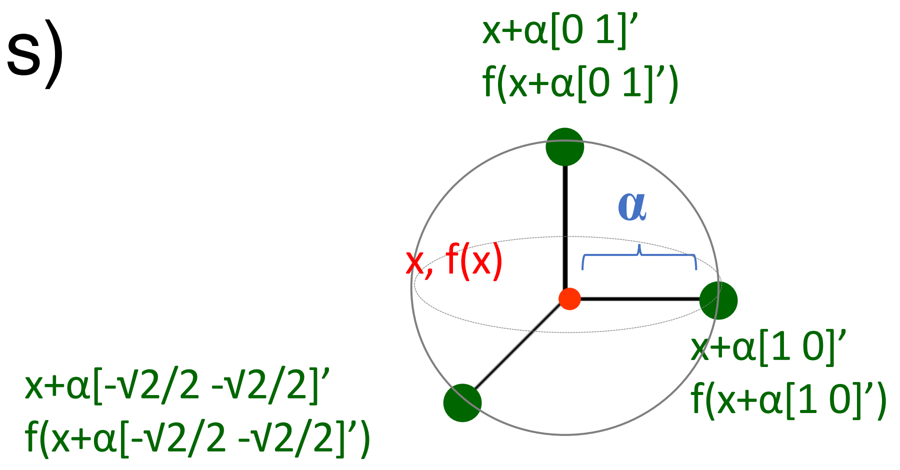
(where *d* is defined by a positive basis)



Linking geometry to gradients

- Assume we sample around a point x on *n+1* points around x in the form

  $x + \alpha d$ where $\alpha > 0$ (where *d* is defined by a positive basis)



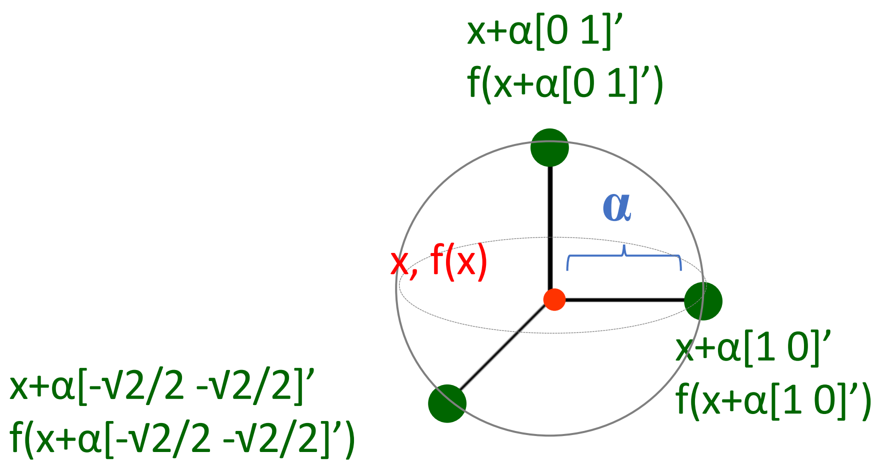
- If sampled objectives at 3 new points is NOT better than the value at x, THEN:
  - The size of the gradient at x **is of the order** of the distance between x and the samples (α) and
  - The order "constant" is a function of the non-linearity of f and the geometry of the sample set
  - As $a \to 0$ … ???

Link to Descent Mechanism

- In order to decrease $f(x)$ it is required $x + \alpha d$ where $\alpha > 0$ to evaluate points on and repeat for decreasing values of α.

- If $\nabla f(x) \neq 0$ (we have not found optimum)

  then there exists α for which: $f(x + \alpha d) < f(x)$



- **Theoretically this scheme is terminated after finite α reductions = Optimization in a finite number of steps**

Summary

- Positive spanning sets are important for guaranteeing descent mechanism is found
- Positive basis sets are formed by linearly independent vectors
- Ensuring using independent vectors minimizes sampling requirements
- Searching on positive spanning sets and reducing step size is linked with magnitude of derivatives in search region