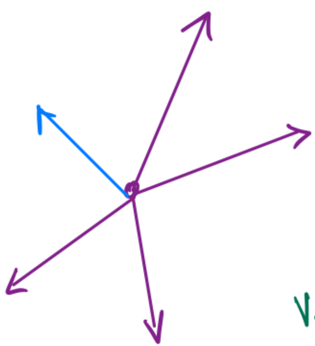
Direct Search Optimization

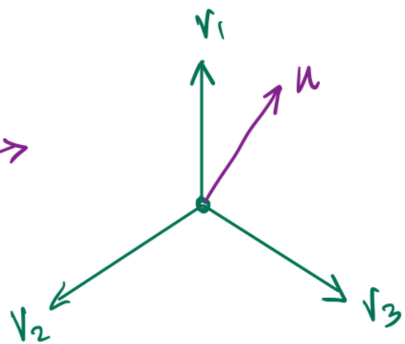
Concept:

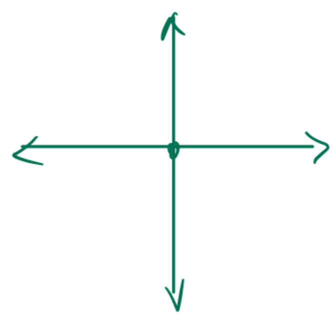
- $\min f(x)$ s.t. $x \in \Omega \subseteq \mathbb{R}^n$
- f may be smooth, possibly not.
- Ω defined by smooth functions
- Improving iterates
- No use of derivative information

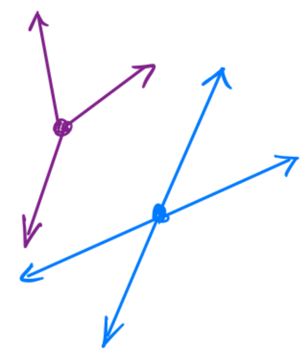
Background Concepts

Def: The positive span of $\beta = \{v_1, v_2, \cdots, v_r\} \subseteq \mathbb{R}^n$ is the convex cone $S = \{x \in \mathbb{R}^n \mid x = a_1v_1 + a_2v_2 + \cdots + a_rv_r,\ a_i \geq 0,\ i = 1, 2, \ldots, r\}$. We say that β is a positive spanning set for S. If no vector $v_i \in \beta$ is in the positive span of $\beta \setminus \{v_i\}$ then β is said to be positively independent. Furthermore, if β is a positively independent positive spanning set for S, then β is a positive basis for S.









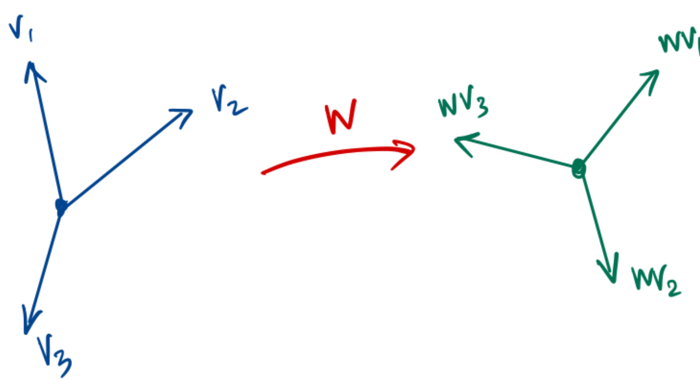
Theorem. Let $\beta = \{v_1, v_2, \ldots, v_r\}$ be a set of nonzero vectors that span $\mathbb{R}^n$. Then β positively spans $\mathbb{R}^n$ if and only if for every nonzero vector $u \in \mathbb{R}^n$, $u^T v_i > 0$ for some $v_i \in \beta$.

Proof. Suppose $\beta = \{v_1, v_2, \ldots, v_r\}$ is a set of nonzero vectors that span $\mathbb{R}^n$.
($\Rightarrow$) Suppose β positively spans $\mathbb{R}^n$. For every nonzero vector $u \in \mathbb{R}^n$ $u = a_1 v_1 + a_2 v_2 + \cdots + a_r v_r$ for some nonnegative coefficients $a_1, a_2, \ldots, a_r$. As $u \neq 0$, $0 < u^T u = (a_1 v_1^T + a_2 v_2^T + \cdots + a_r v_r^T) u = a_1 v_1^T u + a_2 v_2^T u + \cdots + a_r v_r^T u$. Because each $a_i \geq 0$, at least one of $v_i^T u > 0$.

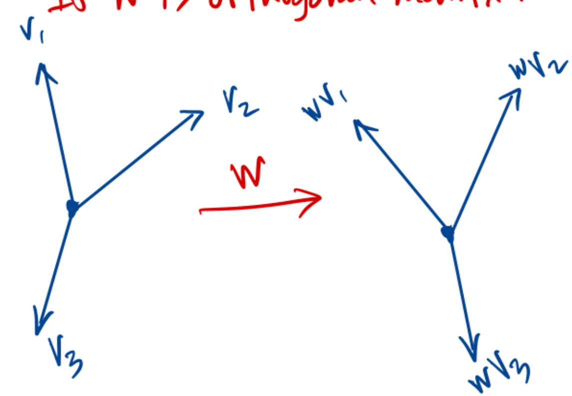
($\Leftarrow$) (contrapositive) Suppose there exists nonzero $u \in \mathbb{R}^n$ such that $v_i^T u \leq 0$ for all i. Use spanning set β to write $u = a_1 v_1 + a_2 v_2 + \cdots + a_r v_r$. Notice that $u^T u = a_1 v_1^T u + a_2 v_2^T u + \cdots + a_r v_r^T u > 0 \Rightarrow a_k < 0$ for some k, that is, β is not a positive spanning set. $\blacksquare$

Theorem. Suppose $\beta = \{v_1, v_2, \ldots, v_r\}$ is a positive basis for $\mathbb{R}^n$ and W a nonsingular $n \times n$ matrix. Then $\gamma = \{Wv_1, Wv_2, \ldots, Wv_n\}$ is also a positive basis for $\mathbb{R}^n$.

Proof. β is a spanning set for $\mathbb{R}^n$. Because W is invertible, γ is a spanning set for $\mathbb{R}^n$. Consider nonzero $u \in \mathbb{R}^n$, then $W^T u \neq 0$ and $(W^T u)^T v_i > 0$ for some index i. So, $u^T(Wv_i) > 0$. Thus, by the previous theorem, γ is also a positive spanning set for $\mathbb{R}^n$. $\blacksquare$



If W is orthogonal matrix?



Corollary. Let I be the $n \times n$ identity matrix, e the vector in $\mathbb{R}^n$ with entries all one, W any nonsingular $n \times n$ matrix. The following matrices consist of columns that form positive bases for $\mathbb{R}^n$.

(a) $\begin{bmatrix} I & -e \end{bmatrix}$

(b) $\begin{bmatrix} I & -I \end{bmatrix}$

(c) $\begin{bmatrix} W & -We \end{bmatrix}$

(d) $\begin{bmatrix} W & -W \end{bmatrix}$

Uniform Angle Positive Basis

we seek a positive basis in $\mathbb{R}^n$ of unit vectors $\{v_1, v_2, \dots, v_{n+1}\}$ satisfying $v_i^T v_j = t$ for all $i \neq j$ and fixed value t.

Consider $w = v_1 + v_2 + \cdots + v_{n+1}$. Then $v_k^T w = 1 + nt$ for each k.

If $1 + nt > 0$ then every v_k lies in the open halfspace $\{x \in \mathbb{R}^n \mid w^T x > 0\}$ and cannot form a positive basis.

Thus $1 + nt < 0$ and $t = -1/n$.

Next, consider matrix $M = \begin{bmatrix} 1 & t & t & \\ t & 1 & t & \\ t & t & 1 & \\ & & & \ddots \\ & & & & 1 \end{bmatrix}$ (which is pos def!)

So we can decompose $M = LL^T \Rightarrow L^T = [v_1 \; v_2 \cdots v_n]$. And finally $v_{n+1} = -\sum_{k=1}^{n} v_k$.

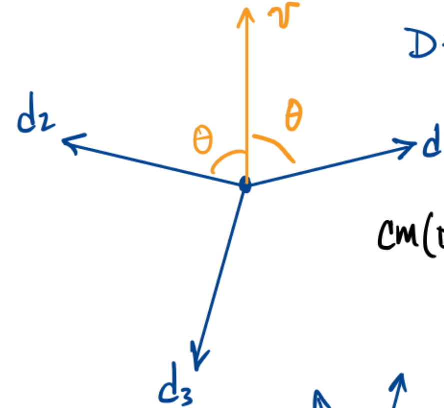
$$v_{n+1}^T v_i = -\sum_{i=1}^{n} v_k^T v_i = \left(\frac{1}{n}\right)(n-1) - 1 = -\frac{1}{n}$$

$$v_{n+1}^T v_{n+1} = -\sum_{i=1}^{n} v_{n+1}^T v_k = -\left(-\frac{1}{n}\right)n = 1$$

Def: The cosine measure of a positive spanning set (of nonzero vectors) D is defined by

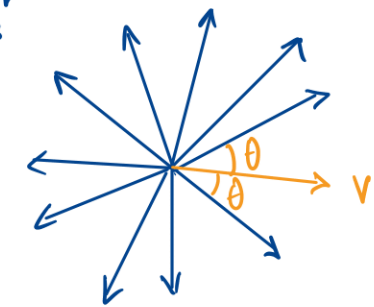
$$\mathrm{cm}(D) = \min_{\substack{v \in \mathbb{R}^n \\ v \neq 0}} \max_{d \in D} \frac{v^T d}{\|v\|\,\|d\|}.$$

$D = \{d_1\ d_2\ d_3\}$

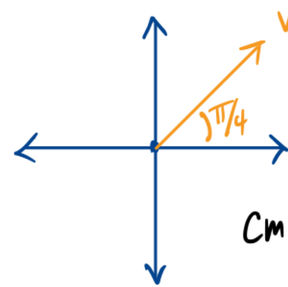


$\mathrm{cm}(D) = \cos\theta$

$$\left(\mathrm{cm}(D) = \max_{\substack{v \in \mathbb{R}^n \\ v \neq 0}} \min_{d \in D} \mathrm{ang}(v, d) \right)$$



Find a vector $v \neq 0$ that maximizes the angle to the closest vector $d \in D$. Then $\mathrm{cm}(D)$ is the cosine of that angle.



$\mathrm{cm}(D) = \cos\frac{\pi}{4} = \frac{\sqrt{2}}{2} \approx 0.71$

$\boxed{0 < \mathrm{cm}(D) < 1}$ for pos. basis. (or pos. spanning set)

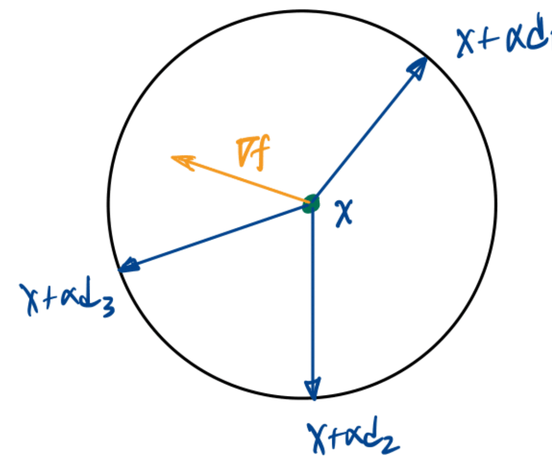
- The cosine measure can be useful in setting bounds on problem-important quantities. For any $v \neq 0$:

$$cm(D) \leq \max_{d \in D} \frac{v^T d}{\|v\| \|d\|}$$

- And then also, there exists a specific vector d satisfying

$$cm(D) \leq \frac{v^T d}{\|v\| \|d\|} = \cos(\text{ang}(v, d))$$

Idea:



Theorem. Let D be a positive spanning set and $\alpha > 0$. Assume ∇f is Lipschitz continuous with constant L in an open set containing $B(x, \Delta)$ where $\Delta = \alpha \max_{d \in D} \|d\|$.

If $f(x) \le f(x+\alpha d)$ for all $d \in D$ then

$$\|\nabla f(x)\| \le \frac{L\Delta}{2\,cm(D)}.$$

Proof: Let $v = -\nabla f(x)$. Then we have

$$cm(D)\,\|\nabla f(x)\|\,\|d\| \le -\nabla f(x)^T d \quad , \text{ and}$$

$$0 \le f(x+\alpha d) - f(x) = \int_0^1 \nabla f(x+t\alpha d)^T (\alpha d)\, dt$$

$$\Rightarrow \quad cm(D)\,\|\nabla f(x)\|\,\|d\|\,\alpha \le \int_0^1 [\nabla f(x+t\alpha d) - \nabla f(x)]\,\alpha d\, dt$$

$$\le \frac{L}{2}\|d\|^2 \alpha^2$$

$$\Rightarrow \quad \|\nabla f(x)\| \le \frac{L\,\|d\|\,\alpha}{2\,cm(D)} \le \frac{L\Delta}{2\,cm(D)}. \qquad \blacksquare$$

Def: The affine hull of set $S \subseteq \mathbb{R}^n$ is the set of all linear combinations of elements of S whose coefficients sum to 1.