

When LP is not a good idea -  
structure in polyhedral optimization  
problems

*Main reference: "Simplicial  
Algorithms for Minimizing Polyhedral  
Functions", CUP 2001.*

M.R.Osborne  
Mathematical Sciences Institute  
Australian National University

**Abstract:** It has been known for 50 years that the discrete  $l_1$  approximation problem can be solved by linear programming (L.P.). However, improved algorithms involve a step which can be interpreted as a line search, and which is not part of the standard simplex algorithm. This is the simplest example of a class of problems with a structure distinctly more complicated than that of the standard nondegenerate LP. Our aim is to uncover this structure for these more general polyhedral functions and to use it to develop what are recognizably simplicial type algorithms.

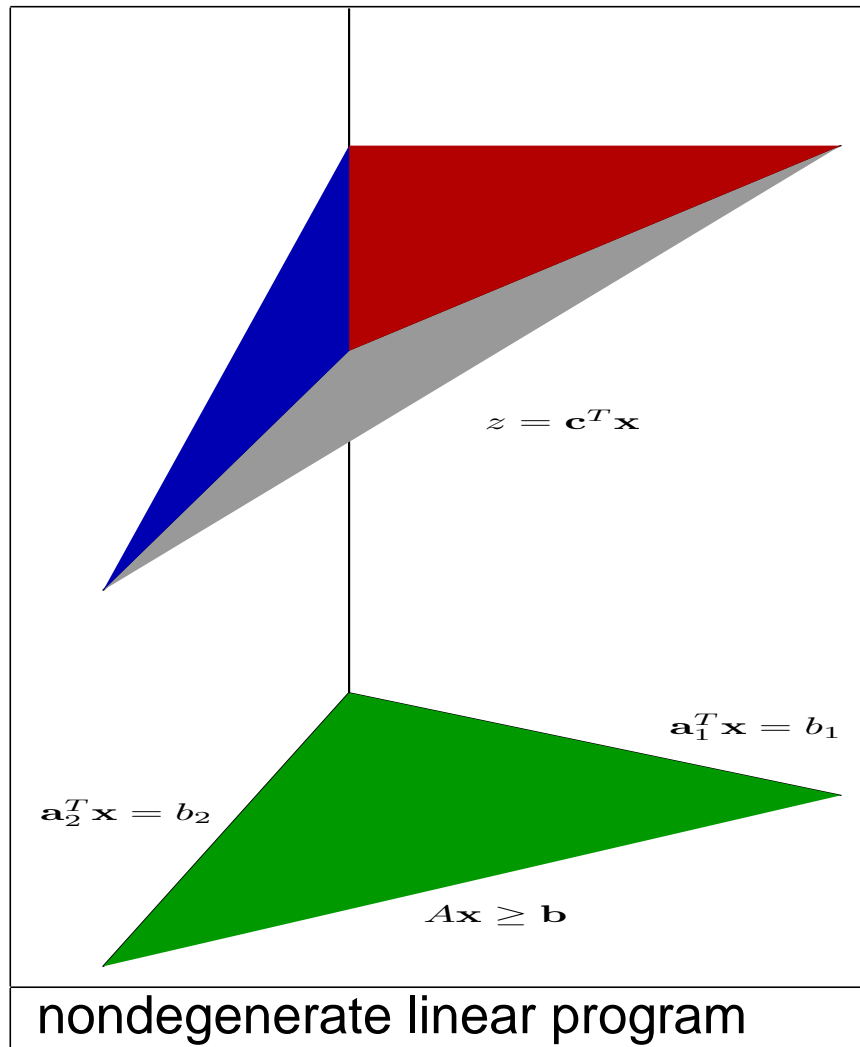
A convex function is the supremum of an affine family:

$$f(\mathbf{x}) = \sup_{i \in \sigma} \mathbf{c}_i^T \mathbf{x} - d_i$$

If the index set  $\sigma$  is finite then  $f(\mathbf{x})$  is polyhedral. The problem of minimizing  $f(\mathbf{x})$  over a polyhedral set  $A\mathbf{x} \geq \mathbf{b}$  can always be written as an LP

$$\min_{A\mathbf{x} \geq \mathbf{b}} h; \quad h \geq \mathbf{c}_i^T \mathbf{x} - d_i, i \in \sigma.$$

*The linear program supports a simple picture!*



Note that three faces of the epigraph intersect at each extreme point  $\mathbf{x} \in R^2$ . The case of degeneracy corresponds here to more than three faces intersecting at an extreme point.

**Problems which arise in discrete estimation** Let the linear model be

$$\mathbf{r} = A\mathbf{x} - \mathbf{b}.$$

Here the estimation problem has the form

$$\min_{\mathbf{x}} F(\mathbf{r}),$$

where  $F(\cdot)$  is a seminorm and polyhedral

We consider algorithms for linear estimation problems which are characterised by:

- 1 The epigraph of  $F(\mathbf{r}(\mathbf{x}))$  is generically degenerate in the sense of linear programming.
- 2 There is a well defined set of necessary conditions which describe the problem optimum and which can be taken here as defining an appropriate sense of nondegeneracy.

It is assumed that  $\text{rank}(A) = p$ , and that this suffices to guarantee a bounded optimum. Associated with extreme points of the epigraph are appropriate sets of algebraic conditions. Typically these involve a subset of the equations specifying the linear model and we refer to this subset as the "active set" at  $\mathbf{x}_\sigma$  where  $\sigma$  is the index set pointing to the active subset.

**Example 1:**  $l_1$  estimation

$$\min_{\mathbf{x}} \sum |r_i| \quad \mathbf{r} = A\mathbf{x} - \mathbf{b} \quad A : R^p \rightarrow R^n.$$

corresponding to

$$\mathbf{c}_j^T = [\pm 1, \pm 1, \dots, \pm 1]A, \quad j = 1, 2, \dots$$

Note apparent redundancy when  $r_i = 0$ . The necessary conditions are:

$$\begin{aligned} 0 &= \sum_{i \in \sigma^C} \theta_i A_{i*} + \sum_{i \in \sigma} u_i A_{i*}, \\ \theta_i &= \text{sign}(r_i), \quad r_i \neq 0, \\ \sigma &= \{i; r_i = 0\}, \\ |u_i| &\leq 1, \quad i \in \sigma. \end{aligned}$$

The nondegeneracy condition is

$$|\sigma| = p, \text{rank}(A_\sigma) = p.$$

In case  $p = 2$  extreme points characterized by (say)

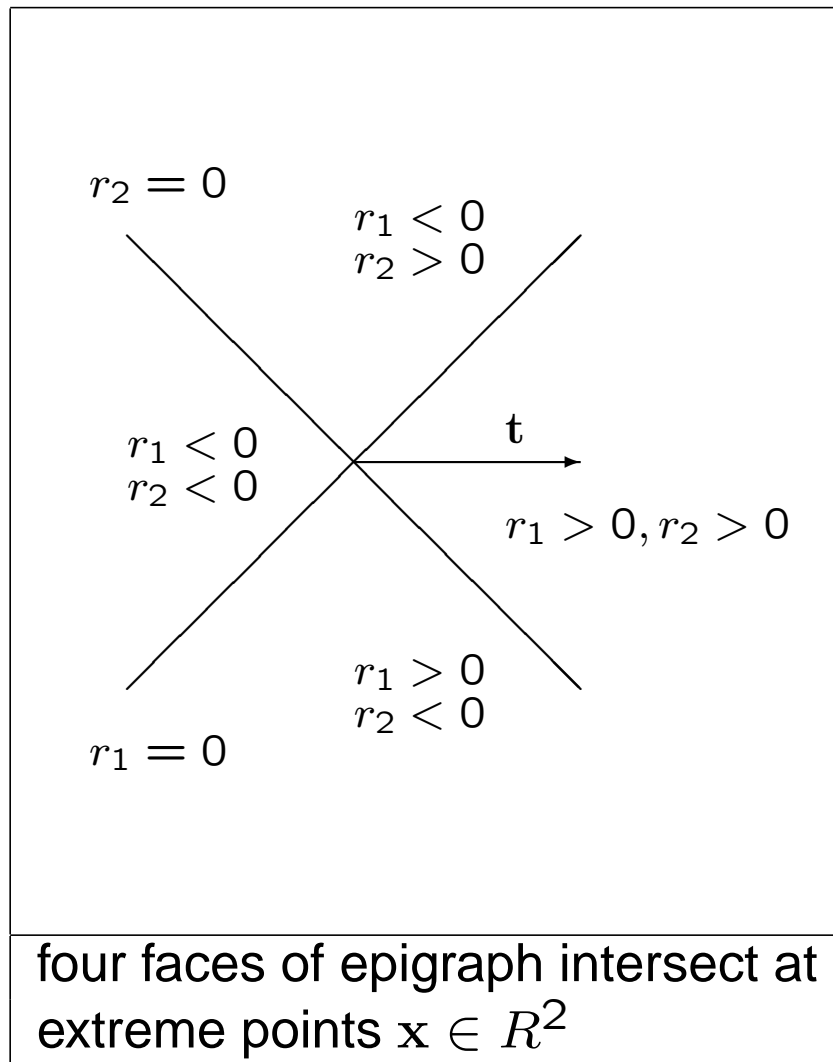
$$\begin{aligned}\pm r_1(x_1, x_2) &= 0, \\ \pm r_2(x_1, x_2) &= 0.\end{aligned}$$

Four faces of epigraph intersect at each extreme point (LP expect 3).  $\pm \Rightarrow \theta_i$  such that directions into faces of epigraph satisfy

$$\begin{aligned}\theta_1 A_{1*} \mathbf{t} &= \lambda_1 > 0, \\ \theta_2 A_{2*} \mathbf{t} &= \lambda_2 > 0.\end{aligned}$$

for convex combination of edge directions. This convention permits each face to be specified unambiguously!





The  $l_1$  problem typically supports a linesearch!

**Example 2:** rank regression.

Let scores  $w_i$  nondecreasing and summing to 0 be given - for example  $w_i = \sqrt{12} \left( \frac{i}{n+1} - \frac{1}{2} \right)$ ,  $i = 1, 2, \dots, n$

$$\min_{\mathbf{x}} \sum_{i=1}^n w_i r_{\nu}(i)$$

$$w_1 \leq w_2 \leq \dots \leq w_n, \quad \sum_{i=1}^n w_i = 0, \quad \|\mathbf{w}\| > 0.$$

$\nu$  ranking set. Nonsmoothness has its origin in the reordering of scores associated with tied residuals. Here the objective is a seminorm.

The necessary conditions are distinctly more complicated!

6 faces of epigraph intersect at extreme points of epigraph over  $R^2$ . Consider the equations characterizing ties:

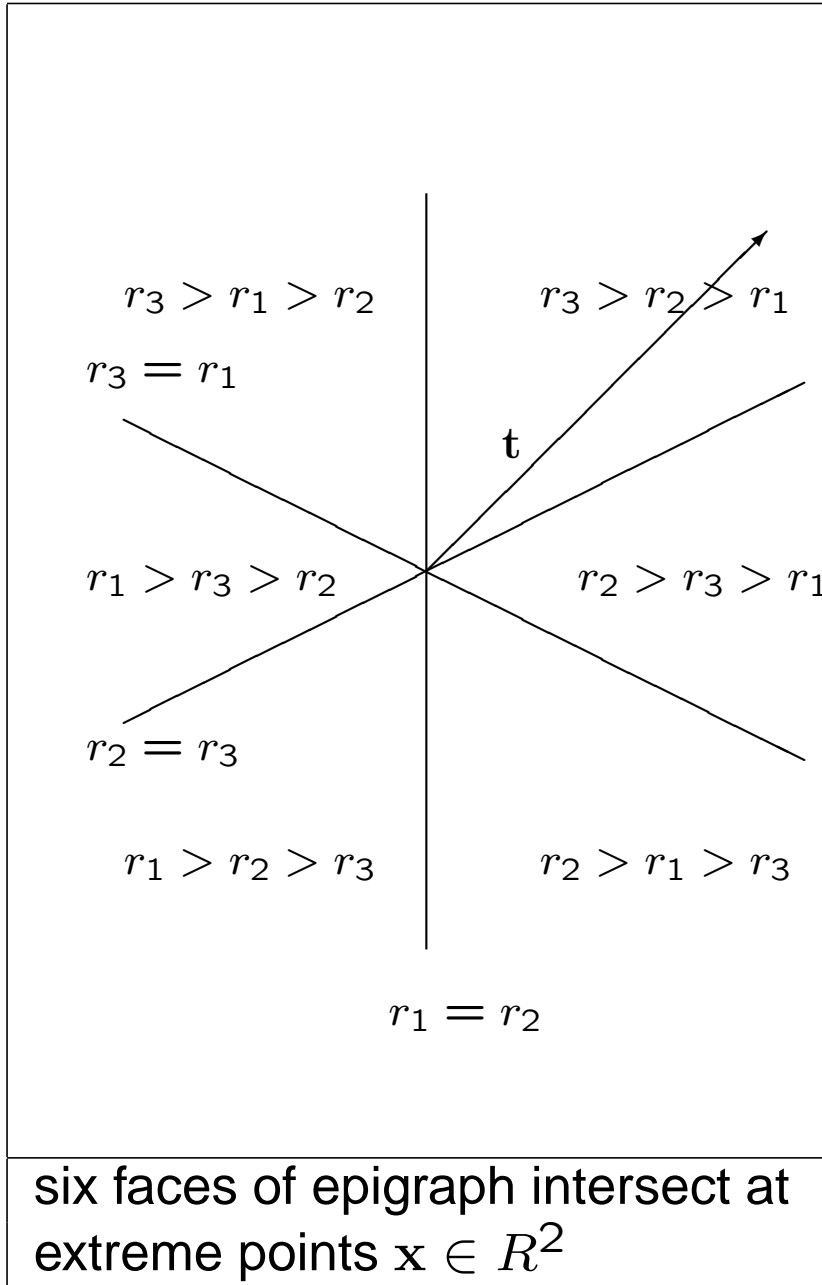
$$\pm (r_2 - r_1) = \pm (r_3 - r_2) = \pm (r_1 - r_3) = 0.$$

Serious redundancy:

$$r_1 - r_3 = -r_3 + r_2 - r_2 + r_1.$$

This implies the third line must pass through intersection of first two. Again redundancy in the association of edges and faces can be resolved by looking at directions into faces as convex combinations of directions along edges.

$$\begin{aligned} \theta_{ik} (A_{i*} - A_{k*}) \mathbf{t} &= \lambda_{ik} > 0, \\ \theta_{kj} (A_{k*} - A_{j*}) \mathbf{t} &= \lambda_{kj} > 0. \end{aligned}$$



## Duality:

$l_1$  Fenchel dual not too bad

$$\min_{\mathbf{u}} \mathbf{b}^T \mathbf{u}, \quad A^T \mathbf{u} = 0, \quad -\mathbf{e} \leq \mathbf{u} \leq \mathbf{e}.$$

**rank regression** Fenchel dual looks familiar, but ...

$$\min_{\mathbf{u}} \mathbf{b}^T \mathbf{u}, \quad A^T \mathbf{u} = 0, \quad \mathbf{u} \in \text{conv} \{ \mathbf{w}_i \}$$

where  $\mathbf{w}_i$  are all distinct permutations of  
 $w_1, w_2, \dots, w_n$ .

$l_1$  is actually a limiting case of rank regression corresponding to sign scores.

## Type 1 PCF:

$$f(\mathbf{x}) = \max_{1 \leq i \leq m} \mathbf{c}_i^T \mathbf{x} - d_i.$$

Set  $\phi_i(\mathbf{r})$ ,  $i = 1, 2, \dots, N$  **structure functionals** for  $f(\mathbf{r}(\mathbf{x}))$  if each extreme point  $\begin{bmatrix} \mathbf{x}^* \\ f(\mathbf{r}(\mathbf{x}^*)) \end{bmatrix}$  of  $\text{epi}(f)$  is determined by the linear system

$$\phi_i(\mathbf{r}(\mathbf{x}^*)) = 0, \quad i \in \sigma \subseteq \{1, 2, \dots, N\}.$$

where  $\sigma$  defines the active set (of structure functionals).

We have already seen examples where the set of structure functionals contains redundant elements!

**Redundancy:** Structure equation  $\phi_s = 0$  is redundant if

$$\exists \pi \neq \emptyset, s \notin \pi \ni (\phi_i = 0 \forall i \in \pi) \Rightarrow \phi_s = 0$$

identically in  $r$ . Consider rank regression example

$$\phi_{12} = r_2 - r_1, \phi_{23} = r_3 - r_2, \phi_{31} = r_1 - r_3.$$

$$\phi_{12} = \phi_{23} = 0 \Rightarrow \phi_{31} = \phi_{23} - \phi_{12} = 0,$$

$$\phi_{12} = 0 \Rightarrow \phi_{21} = 0.$$

multiplication by -1 is significant!

$\phi_{12}$ ,  $\phi_{23}$  and  $\phi_{23}$ ,  $\phi_{31}$  examples of nonredundant pairs.

Say: get nonredundant configurations by *allowable reductions*.

**Linear independence:** Given set of structure functionals

$$\text{rank}(V_\sigma) = k = |\sigma| \leq p.$$

$$V_\sigma^T = \Phi_\sigma^T A \in R^p \rightarrow R^k$$

$$\Phi_\sigma = \left[ \nabla_r \phi_{\sigma(1)}^T \quad \cdots \quad \nabla_r \phi_{\sigma(k)}^T \right] \in R^k \rightarrow R^n.$$

**Nondegeneracy:** Each allowable reduction of active set is linearly independent.

To generate a compact local representation specialise one of the allowable reductions. Let  $\mathbf{x} = \mathbf{x}^* + \varepsilon \mathbf{t}$ ,  $\varepsilon > 0$  small enough. Then, using piecewise linearity of the objective, rearranging gives

$$f(\mathbf{r}(\mathbf{x})) = f_\sigma(\mathbf{r}(\mathbf{x})) + \sum_{i=1}^{|\sigma|} \omega_i(\mathbf{t}) \phi_{\sigma(i)}(\mathbf{r}(\mathbf{x})),$$

1.  $f_\sigma$  smooth,  $\omega_i(\mathbf{t})$  provides nonsmooth behaviour.
2. Each distinct realization of  $\omega_i(\mathbf{t})$ ,  $i = 1, 2, \dots, |\sigma|$  characterizes one of the faces of  $\text{epi}(f)$  meeting at  $\begin{bmatrix} \mathbf{x}^* \\ f(\mathbf{r}(\mathbf{x}^*)) \end{bmatrix}$ .



An alternative is to use the property displayed in the examples to develop a **local** description by characterizing the individual faces at a particular extreme point (more generally a nonsmooth point of the epigraph).

**Completeness:** For each face  $s$ ,  $1 \leq s \leq q$  of  $\mathcal{T}(\text{epi}(f), \mathbf{x}^*)$  there exists  $\sigma_s$  such that directions into the face

$$\begin{bmatrix} \mathbf{x}^* + \varepsilon \mathbf{t} \\ f(\mathbf{x}^* + \varepsilon \mathbf{t}) \end{bmatrix} = \begin{bmatrix} \mathbf{x}^* \\ f(\mathbf{x}^*) \end{bmatrix} + \varepsilon \begin{bmatrix} \mathbf{t} \\ f'(\mathbf{x}^* : \mathbf{t}) \end{bmatrix}$$

satisfy

$$V_{\sigma_s}^T \mathbf{t} = \lambda > 0.$$

Note that  $\forall s$  system

$$\phi_{\sigma_s(i)}(\mathbf{x}) = 0, \quad i = 1, 2, \dots, p$$

has same solution  $\mathbf{x}^*$ .

Redundancy is an algebraic property. At an extreme point it requires  $q > p + 1$ .

### **Extreme directions in $\mathcal{T}(\text{epi}(f), \mathbf{x}^*)$**

These are given by

$$\mathbf{t}_{\sigma_s}^i = V_{\sigma_s}^{-T} \mathbf{e}_i, \quad i = 1, 2, \dots, p, \quad s = 1, 2, \dots, q.$$

Edges formed by the intersection of adjacent faces (say  $\sigma_s, \sigma_t$ ) are determined by an equation of this form for each face and there is potential here for overspecification.

*What characterizes an edge unambiguously is that a particular structure functional in the allowable reductions increases away from zero.*

**Example  $l_1$  estimation:** Active structure functionals correspond to zero residuals.

$$\phi_{2i-1} = r_i, \phi_{2i} = -r_i, N = 2n.$$

Let an extreme point  $x^*$  be determined by

$$\phi_{\sigma(i)} = r_i, \sigma = \{1, 3, \dots, 2p - 1\}$$

Let  $x = x^* + \varepsilon t$ . Then

$$f(\mathbf{r}(\mathbf{x})) = \sum_{|r_i(\mathbf{x}^*)| > 0} |r_i| + \sum_{i=1}^p \omega_i(\mathbf{t}) \phi_{\sigma(i)}(\mathbf{r}(\mathbf{x}))$$

For each allowable reduction of the active structure functionals  $\omega_i(\mathbf{t}) \phi_{\sigma(i)}(r_i) = |r_i|$ ,  $\omega_i = \pm 1$ .

Completeness needs finer structure. For face

$$r_1 > 0, r_2 > 0, r_3 > 0 : \sigma_s = \{1, 3, 5\}$$

$$r_1 > 0, r_2 < 0, r_3 > 0 : \sigma_s = \{1, 4, 5\}$$

Differences between sets of equations for extreme directions are pretty trivial in this case.

There are  $2^p$  faces intersecting at  $x^*$ .

**Example: rank regression: Structure is in ties**

$$\phi_{ij} = r_j - r_i, 1 \leq i \neq j \leq n, N = n(n - 1).$$

Redundancies

$$\phi_{ij} = -\phi_{ji}, \phi_{ik} = \phi_{jk} + \phi_{ij}.$$

Possible structure equations when  $p = 3$ .

$$r_1 = r_2 = r_3 = r_4,$$

$$r_1 = r_2, r_3 = r_4 = r_5,$$

$$r_1 = r_2, r_3 = r_4, r_5 = r_6.$$

In first case  $\sigma_1 = \{\phi_{12}, \phi_{13}, \phi_{14}\}$  possible set of structure functionals - specializes  $r_1$  (origin!).

$$f(\mathbf{r}) = \sum_{i=5}^n w_{\mu(i)} r_i + \left( \sum_{i=l}^{l+4} w_i \right) r_1 + \sum_{i=2}^4 \omega_{i-1}(\mathbf{t}) \phi_{1i}.$$

## Redundancy v's completeness!

This set is not good for completeness. If t is into face

$$r_1 < r_2 < r_3 < r_4$$

$$r_1 < r_2 < r_3 < r_4 \Rightarrow \phi_{12} > 0, \phi_{13} > \phi_{12}, \phi_{14} > \phi_{13}.$$

Relaxed structure functionals do not give right ordering.

Right set is ( $\sigma_s = \{12, 23, 34\}$  )

$$\phi_{12} > 0, \phi_{23} = \phi_{13} - \phi_{12} > 0, \phi_{34} = \phi_{14} - \phi_{13} > 0.$$

*Changing structure functional basis changes representation of non-smooth part of function.*

$$\begin{aligned}
 \sum_{i=1}^p \omega_i(\mathbf{t}) \phi_{\sigma_1(i)}(\mathbf{x} + \mathbf{t}) &= \mathbf{t}^T V_{\sigma_1} \boldsymbol{\omega}(\mathbf{t}), \\
 &= \mathbf{t}^T V_{\sigma_1} S_s S_s^{-1} \boldsymbol{\omega}(\mathbf{t}), \\
 &= \sum_{i=1}^p (\omega_s(\mathbf{t}))_i \phi_{\sigma_s(i)}(\mathbf{x} + \mathbf{t})
 \end{aligned}$$

where  $\phi_{\sigma_1}^T S_s = \phi_{\sigma_s}^T$

$$\begin{aligned}
 &\begin{bmatrix} \phi_{12} & \phi_{13} & \phi_{14} \end{bmatrix} \begin{bmatrix} 1 & -1 & \\ & 1 & -1 \\ & & 1 \end{bmatrix} \\
 &= \begin{bmatrix} \phi_{12} & \phi_{23} & \phi_{34} \end{bmatrix}
 \end{aligned}$$

Solutions of systems  $V_{\sigma_s}^T \mathbf{t}_i^s = \mathbf{e}_i$ ,  $i = 1, 2, 3$ , break ties

$$\mathbf{t}_1^s : r_1 < r_2 = r_3 = r_4,$$

$$\mathbf{t}_2^s : r_1 = r_2 < r_3 = r_4,$$

$$\mathbf{t}_3^s : r_1 = r_2 = r_3 < r_4.$$

**subdifferential:** Let  $f(\mathbf{x})$ ,  $\mathbf{x} \in X$  be convex. The subdifferential  $\partial f(\mathbf{x})$  is the set

$$\{\mathbf{v}; f(\mathbf{t}) \geq f(\mathbf{x}) + \mathbf{v}^T (\mathbf{t} - \mathbf{x}), \forall \mathbf{t} \in X\}.$$

Also subdifferential is convex hull of gradient vectors at nearby differentiable points.

Subgradient  $\mathbf{v}$  generalises idea of a gradient vector at points of nondifferentiability of  $f(\mathbf{x})$ .

Subdifferential is important for characterizing optima and calculating descent directions in nonsmooth convex optimization.

**directional derivative:**

$$\begin{aligned} f'(\mathbf{x} : \mathbf{t}) &= \inf_{\lambda > 0} \frac{f(\mathbf{x} + \lambda \mathbf{t}) - f(\mathbf{x})}{\lambda}, \\ &= \max_{\mathbf{v} \in \partial f(\mathbf{x})} \mathbf{v}^T \mathbf{t}. \end{aligned}$$

**optimality:**  $\mathbf{x}$  minimizes  $f(\mathbf{x})$  if  $0 \in \partial f(\mathbf{x})$ .

Recall

$$f(\mathbf{r}(\mathbf{x})) = f_\sigma(\mathbf{r}(\mathbf{x})) + \sum_{i=1}^{|\sigma|} \omega_i(\mathbf{t}) \phi_{\sigma(i)}(\mathbf{r}(\mathbf{x})),$$

Implies a representation of subdifferential:

$$\mathbf{v}^T \in \partial f(\mathbf{r}(\mathbf{x})) \rightarrow \mathbf{v} = \mathbf{f}_g + V_\sigma \mathbf{z}.$$

$$\mathbf{f}_g = \nabla_x f_\sigma(\mathbf{r})^T : \text{gradient of smooth part.}$$

$$(V_\sigma)_{*i} = \nabla_x \phi_{\sigma(i)}^T = \left\{ \nabla_r \phi_{\sigma(i)} A \right\}^T, i = 1, 2, \dots, |\sigma|,$$

$$\mathbf{z} \in Z_\sigma = \text{conv}(\omega_s, s = 1, 2, \dots, q).$$

Standard inequality for directional derivative gives

$$Z_\sigma = \left\{ \mathbf{z}; (\mathbf{f}_g + V_\sigma \mathbf{z})^T \mathbf{t} \leq f'(\mathbf{x}^* : \mathbf{t}) \right\}.$$

Constraint set known if directional derivative known.



## Role of extreme directions

Extreme points of  $Z_\sigma$  are determined by the extreme directions associated with the edges of  $\mathcal{T}$  ( $\text{epi}(f), \mathbf{x}^*$ ).

Key calculation is

$$\begin{aligned} f'(\mathbf{x}^* : \mathbf{t}_s) &= \mathbf{f}_g^T \mathbf{t}_s + \max_{\mathbf{z} \in Z_\sigma} \mathbf{z}^T V_\sigma^T \mathbf{t}_s, \\ &= \mathbf{f}_g^T \mathbf{t}_s + \max_{\mathbf{z} \in Z_\sigma} \left\{ \sum_{i=1}^p \lambda_i \mathbf{z}^T V_\sigma^T \mathbf{t}_i^s \right\}, \\ &\leq \mathbf{f}_g^T \mathbf{t}_s + \sum_{i=1}^p \lambda_i \max_{\mathbf{z} \in Z_\sigma} \mathbf{z}^T V_\sigma^T \mathbf{t}_i^s \\ &= \sum_{i=1}^p \lambda_i f'(\mathbf{x}^* : \mathbf{t}_i^s). \end{aligned}$$

It uses the linearity of  $f(\mathbf{x})$  on the faces of  $\mathcal{T}$  twice.

## Computation of $Z_\sigma$ :

Start with tie  $r_1 = r_2 = r_3 = r_4$ . If  $r_1$  leaves group on edge then  $\phi_{12}, \phi_{13}, \phi_{14}$  all relax. On the edge  $\phi_{23} = \phi_{24} = 0$  have to relate  $\phi_{12}, \phi_{13}, \phi_{14}$  and  $\phi_{12}, \phi_{23}, \phi_{24}$ . In general

$$\begin{bmatrix} \Phi_j & \nabla_r \phi_j^T \end{bmatrix} \begin{bmatrix} S_j & \\ \mathbf{s}_j^T & 1 \end{bmatrix} = \Phi_\sigma P_j.$$

where active set condition on edge is  $\Phi_j^T A \mathbf{t} = 0$ .

$$\begin{aligned} f'(\mathbf{x}^* : \mathbf{t}) &= \mathbf{f}_g^T \mathbf{t} + \max_{\mathbf{z} \in Z_\sigma} \mathbf{z}^T V_\sigma^T \mathbf{t}, \\ &= \mathbf{f}_g^T \mathbf{t} + \max_{\mathbf{z} \in Z_\sigma} \mathbf{z}^T P_j^{-T} \begin{bmatrix} S_j^T & \mathbf{s}_j \\ & 1 \end{bmatrix} \begin{bmatrix} \Phi_j^T \\ \nabla_r \phi_j \end{bmatrix} A \mathbf{t}, \\ &= \mathbf{f}_g^T \mathbf{t} + \max_{\mathbf{z} \in Z_\sigma} \mathbf{z}^T P_j^{-T} \begin{bmatrix} \mathbf{s}_j \\ 1 \end{bmatrix} \mathbf{v}_j^T \mathbf{t}, \\ &= \mathbf{f}_g^T \mathbf{t} + \begin{cases} \zeta_j^+ \mathbf{v}_j^T \mathbf{t}, & \mathbf{v}_j^T \mathbf{t} > 0, \\ \zeta_j^- \mathbf{v}_j^T \mathbf{t}, & \mathbf{v}_j^T \mathbf{t} < 0. \end{cases} \end{aligned}$$

This gives the inequalities determining  $Z_\sigma$  in the form

$$\zeta_j^- \leq \begin{bmatrix} \mathbf{s}_j^T & 1 \end{bmatrix} \mathbf{z} \leq \zeta_j^+.$$

Result for rank regression is that for each group of ties in the active set

$$\left( \sum_{i=1}^{m-k+1} w_i \right) \leq \sum_{i=k}^m z_{\pi(i)} \leq \left( \sum_{i=k+1}^{m+1} w_i \right).$$

Here  $\pi$  is the set of all permutations of  $m - k + 1$  components of the multiplier vector,  $k = 1, 2, \dots, m$ , and  $m + 1$  is the number of residuals involved in current group.

If  $0 \notin \mathbf{g} + V\mathbf{z}$ ,  $\mathbf{z} \in Z$  then most violated multiplier condition can be used to generate a descent direction corresponding to relaxing off an active structure functional - essentially a standard "simplex" step.

**Linesearch:** Descent step relaxes off one structure functional while remainder defining edge stay active (simplicial step). Experience is that linesearch in this direction is profitable. Linesearch must terminate at new active structure functional.

$l_1$ : Only necessary to know distances to nonsmooth points in search direction. Required point is a weighted median. Hoare's partitioning algorithm suggested with partition bound defined by standard median of three.

General: Bisection applied to directional derivative to refine bracket of minimum. Explicit computation when bracket contains just one active member.

Statistical estimation: Asymptotic linearity results suggest use of secant algorithm to find axis crossing point of piecewise constant directional derivative. Shifting strategy important.

These are all partitioning methods. Important that evaluation of  $f'(\mathbf{x} : t)$  is no worse than  $n\gamma(n)$ ,  $\gamma(n)$  of slow growth.

