Inequality constraints

- Assume a non-linear objective
- write constraints

$$\mathcal{I} = \{ \text{inequalities}, \ a_i^T x \geq b_i \}$$

$$\mathcal{E} = \{ \text{equalities}, \ a_i^T x = b_i \}$$

- objective

$$\min \ x^T \frac{G}{2} x + d^T x$$

notice that
G pd $\Rightarrow$ convex (easyish)
otherwise HARD

- Lagrangian:

$$\mathcal{L} = x^T \frac{G}{2} x + d^T x - \sum_{i \in \mathcal{I} \cup \mathcal{E}} \lambda_i (a_i^T x - b_i)$$

- Define Active set

$$A(x^*) = \{ i \in \mathcal{E} \cup \mathcal{I} \mid a_i^T x^* = b_i \}$$

constraints that are ACTIVE at soln.

this gives as KKT

$$Gx^* + b - \sum_{i \in A(x^*)} \lambda_i a_i = 0$$

$$a_i^T x^* = b_i \qquad i \in A(x^*)$$

$$a_i^T x^* \geq b_i \qquad \cancel{\text{all othe}}\ i \in \mathcal{I} - A(x^*)$$

$$\lambda_i \geq 0 \qquad i \in \mathcal{I} \cap A(x^*)$$

Notice: if we know the active set, we could get x^* easily from a linear system (but we don't)

Strategy:

- start with a feasible point
- improve, while keeping it feasible
- maintain model of $A(x^*)$
  W_k = working set.

Taking a step

$$p = x - x_K$$
$$g_K = G x_K + d.$$

so we must min

$$\frac{p^T G p}{2} + g_K^T p + \left[\text{constant depending on } x_K \right]$$

problem:

$$\min \quad \frac{p^T G p}{2} + g_K^T p$$
$$\text{st} \quad a_i^T p = 0 \quad , \quad i \in W_K$$

Solve to get p_K, which if feasible for W_K (but may not be for others).

now $a_i^T (x_K + \alpha p_K) = b_i$ so

$x_K + \alpha p_K$ is feasible for W_K

<u>Now</u>, p_k could be feasible for <u>all</u> constraints

$$\rightarrow x_{k+1} = x_k + p_k.$$

- otherwise, search $\alpha \in [0,1)$ to find an α so all constraints are satisfied

  - consider $i \in \mathcal{I} - W_k$

    - if $a_i^T p_k \geq 0$ then

      $$a_i^T (x_k + \alpha p_k) \geq a_i^T x_k \geq b_i$$

      so <u>any</u> $\alpha \in [0,1)$ is OK

    - if $a_i^T p_k < 0$ then

      $$\alpha_k \leq \frac{b_i - a_i^T x_k}{a_i^T p_k}$$

- walk constraints to find ~~smallest~~ <u>largest</u> α_k

constraints for which α_k is min are <u>blocking</u> constraints (α_k could even be zero - then a blocking constraint isn't in W_k)

- Now insert a blocking constraint into W_k to get W_{k+1}

<u>Overall</u>

- Start x_0 feasible
- Until finished
  - Until x_k is minimizer of Quad prog over W_k
    - Take a step, as above
    - minimizer when $p = 0$
  - Any -ve lagrange mults?
    - Yes? remove one from WS.
    - No? finished

(Housekeeping: show that step after dropped constraint is feasible).

Issues:

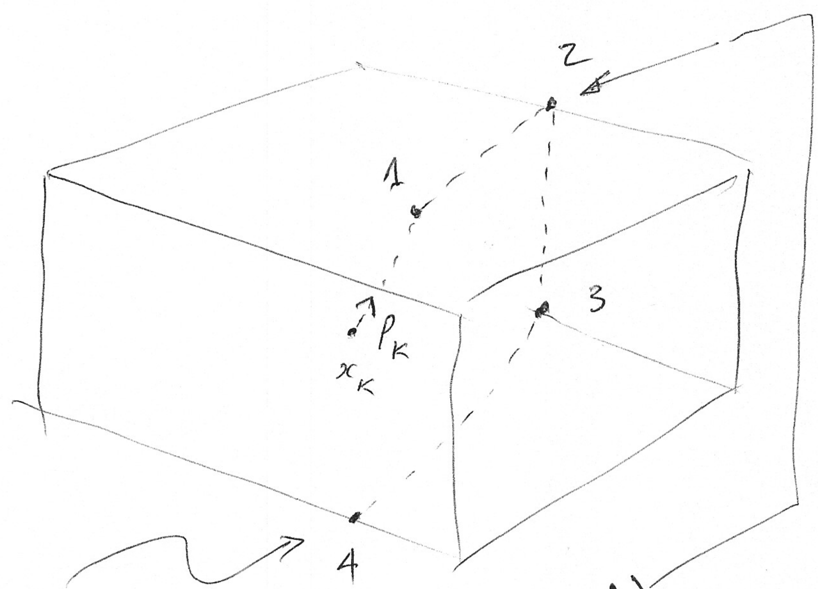
- if G not pd, this won't work
- in fact, difficult points for Any active set alg exist
- Not good for large problems, because we go through constraints slowly.

Large scale active set methods:

- "Box" constraints are special

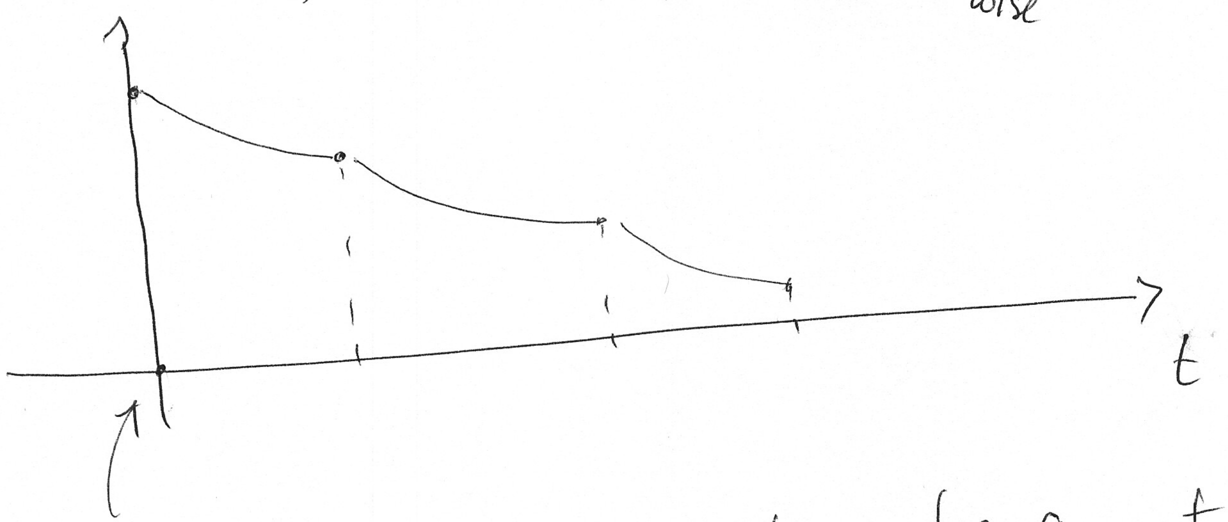
$$l \leq x \leq u$$

assume feasible x_k, search dirn p_k.



$w(\ ) =$ Wrapped around box

consider $f(w(x_k + t p_k))$. Cont. Piecewise Quadratic



<u>Strategy</u>: · search along t for first local min

- Cauchy point x^c is first local minimizer.
- find by searching segments (easy).
- Now at Cauchy point, generate improvement, looking at ~~bound~~ blocking constraints constraints

- e.g.:

$$x^T G x + d^T x$$

$$x_i = x_i^c$$ ← ~~bou~~ active; we are on this face of box

$$l \leq x \leq u$$

- we can rewrite, plugging in values sothat

$$x = \begin{pmatrix} 0 \\ \vdots \\ x_i \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} v_1 \\ \vdots \\ 0 \\ \\ v_N \end{pmatrix}$$ ← unknowns

↑ active

<u>This gets</u>

$$v^T M v + w^T v$$

$$\text{st}\quad \hat{l} \le v \le \hat{u}$$

- now do C.G., starting w/ $v = \begin{pmatrix}\text{relevant} \\ \text{elements} \\ \text{of} \\ x^c\end{pmatrix}$

  and stopping one step before constraint violation.

- This gets new points.

Common application: in important cases, one may be able to write the dual directly.

SVM

$$\min \quad \frac{w'w}{2}$$

$$\text{st} \quad y_i\,(w'x_i + b) \geq 1$$

Primal form, Separable

$$\mathcal{L}_P(w,\lambda) = \frac{w'w}{2} - \sum_i \lambda_i \left\{ \left[y_i\,(w'x_i + b) \right] - 1 \right\}$$

$$\nabla_w \mathcal{L} = 0 = w - \sum_i \lambda_i \left\{ \left[y_i x_i \right] \right\}$$

$$\nabla_b \mathcal{L} = 0 = -\sum_i \lambda_i y_i$$

Subst

$$\mathcal{L}_D = \sum_i \lambda_i - \frac{1}{2} \sum_{ij} \lambda_i \lambda_j \left[y_i y_j (x_i^T x_j) \right]$$

Notice constraints

$$\sum_i \lambda_i y_i = 0$$

$$\lambda_i \geq 0$$

and we must max this in λ

If there is an fp for primal, the max is soln to primal

i.e. Value (Dual) = Value (Primal)