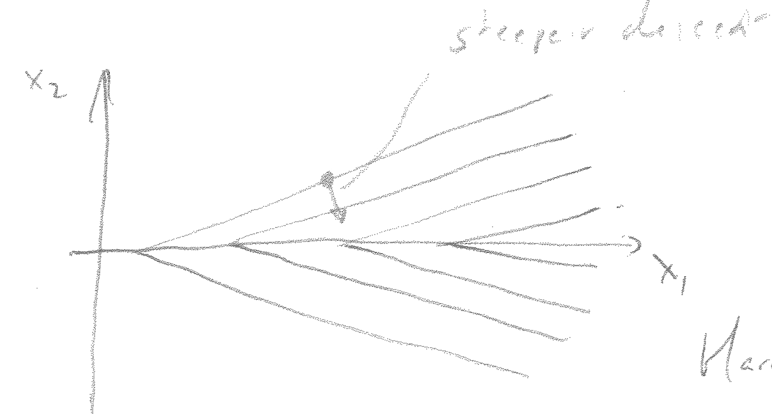
Nonsmooth Optimization

Gradient functions have discontinuities. Eg. l_1-penalty function.

Eg: $\min \ -x_1 + 100|x_2|$ s.t. $0 \le x_1 \le 10$.

Optimal soln: $x_1 = 10$, $x_2 = 0$.

Contours:

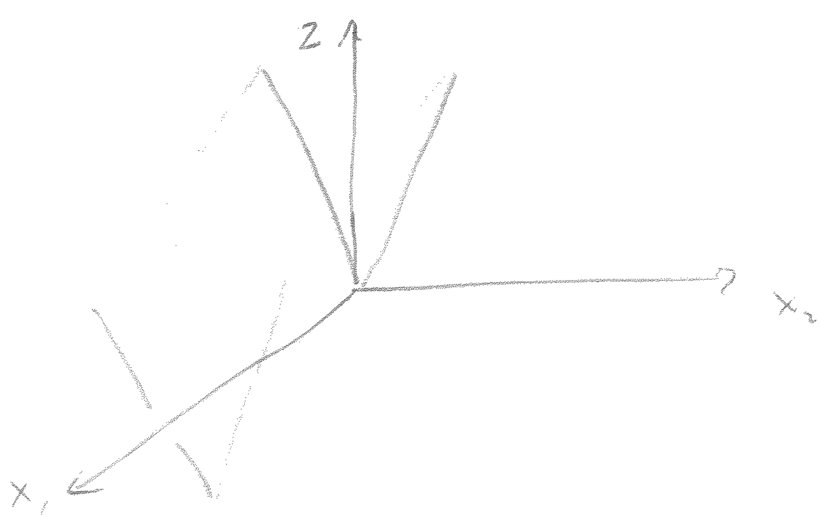


Hard to choose good direction.

Hard to ~~choose~~ find good step length algebraically.

$$\nabla f = \begin{cases} \begin{bmatrix} -1 \\ +100 \end{bmatrix} & \text{if } x_2 > 0 \\ \begin{bmatrix} -1 \\ -100 \end{bmatrix} & \text{if } x_2 < 0 \end{cases}$$

Graph of function:



More involved example:

$$\max_{y,S} \left(b^T y + \lambda_{\min}(S)\right)$$
s.t. $\sum A_i y_i + S = C$

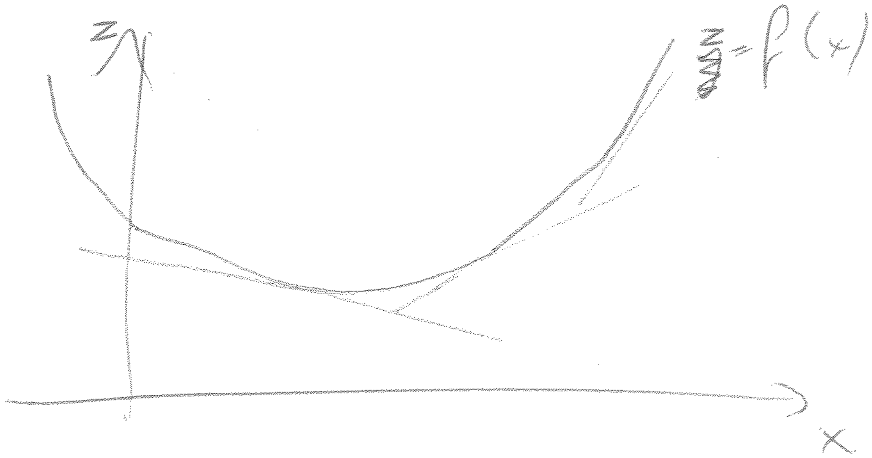
A_i, C, S square symmetric matrices.

y vector

$$\min \; f(x) \qquad \text{(NSP)}$$
$$\text{s.t.} \; g_i(x) \le 0 \quad i=1,\dots,m.$$

Assume functions are convex, ~~both~~ continuous, but may not be differentiable

Can approximate using LP:



~~At a point $\bar{x}$:~~

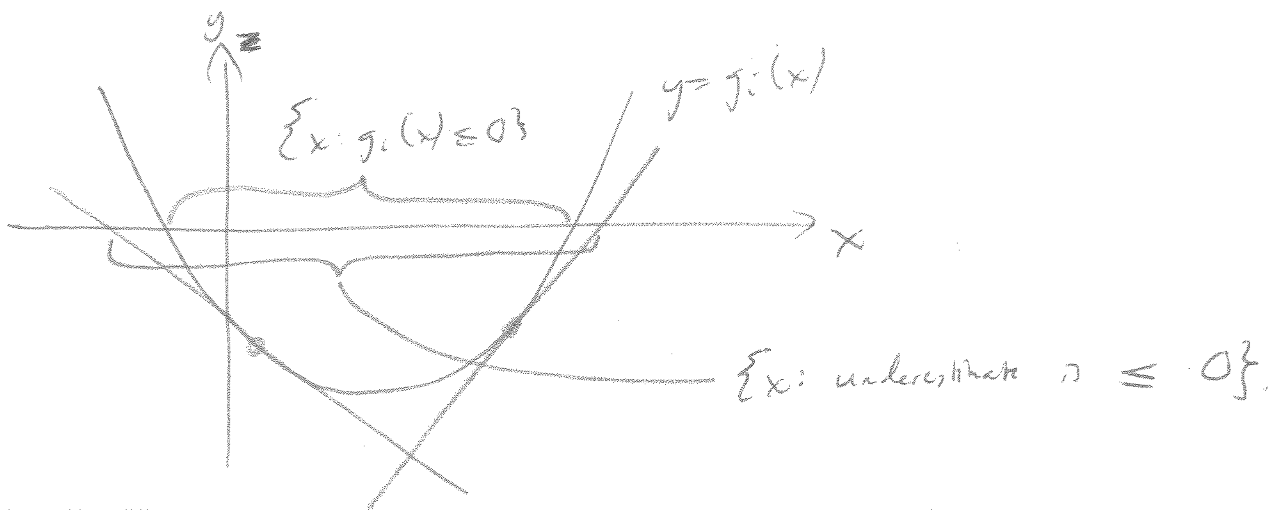
Rewrite problem as

$$\min \; z$$
$$\text{s.t.} \; f(x) - z \le 0$$
$$g_i(x) \le 0 \quad i=1,\dots,m$$

Approximate constraints from underneath: (as with Kelley's cutting plane algorithm)

$$\cancel{f(x)} \; g_i(x) \ge g_i(\bar{x}) + \xi^T(x-\bar{x})$$

ξ is a subgradient of g_i at $\bar{x}$.



Can underestimate $f(x) - z$ similarly.

So get LP:

$$\text{(LP)} \quad \begin{array}{ll} \min & z \\ \text{s.t.} & Ax + hz \le b \end{array}$$

relaxation of (NSP)

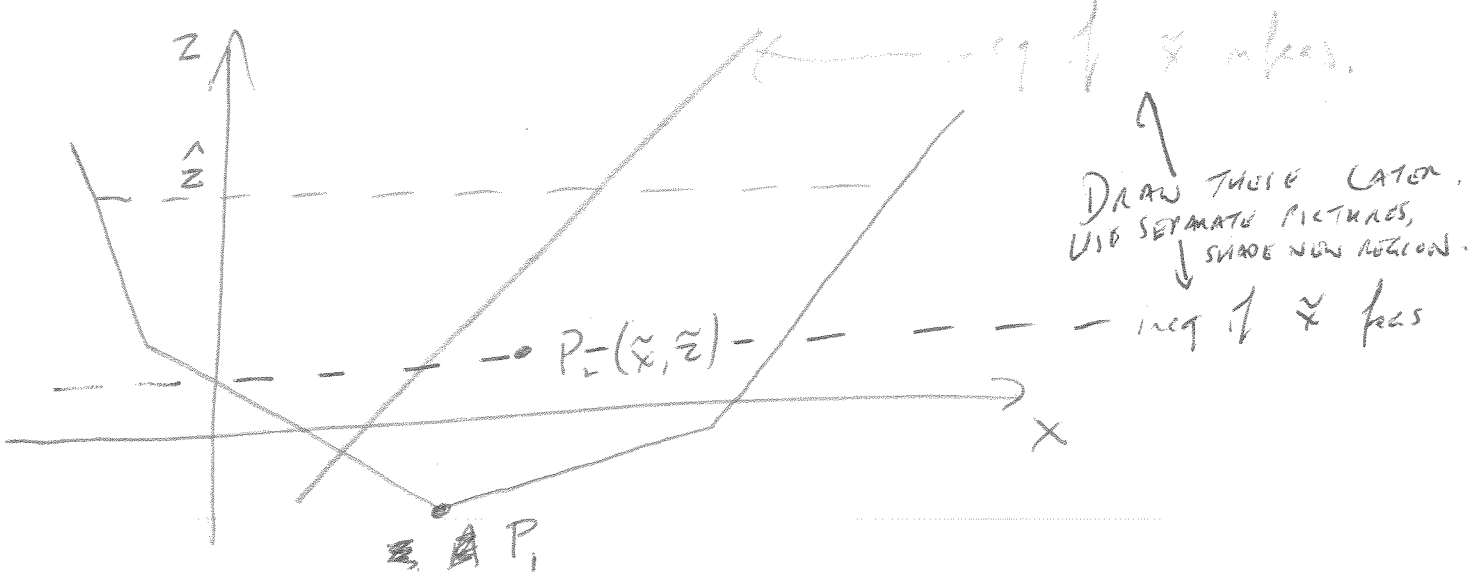
for appropriate A, h, b, generated from supporting inequalities.

Solve (LP) using an algorithm for linear programming. Can use simplex, or an interior point method.

Note that any feasible point $\hat{x}$ gives an upper bound $\hat{z} = f(\hat{x})$ on optimal value of (NSP).

If use an interior point method, best not to solve problem exactly:

Alec 107



If solve ~~exactly~~ (LP) exactly, get (P_1). But hard to ~~solve~~ restart an interior point method from an extreme point.
So find analytic center (or a good point in the middle), P_2

Let P_2 be $(\tilde{x}, \tilde{z})$.

~~If $\tilde{x}$ infeasible, add c~~
If $\tilde{x}$ violates $g_i(x) \le 0$, add supporting ineq for this constraint and resolve:

If $\tilde{x}$ feasible, add constraint $z \le \tilde{z}$ and find new center

Algo finds a solution within ε of optimality in $O(n/\varepsilon)$ iterations.

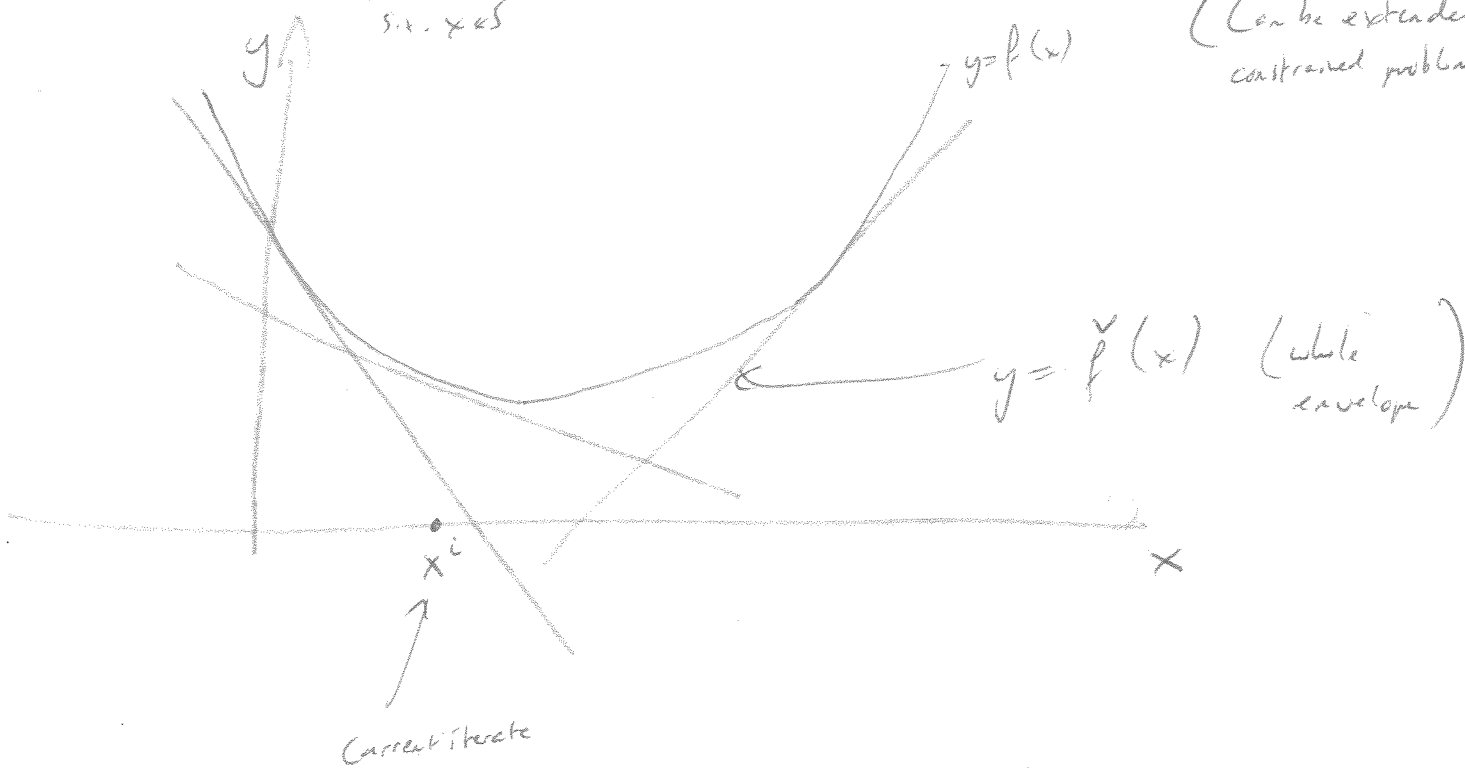
If also allow dropping of constraints, can get within ε of optimality in $O(n \ln(\frac{1}{\varepsilon}))$ iterations.

Only need a handful of iterations to find new center after adding a cut
(can add multiple cuts at once.

Bundle Methods

$\min_{\text{s.t. } x \in S} f(x)$ where f is convex, continuous, nonsmooth.

(Can be extended to constrained problems.)



Have upper bound: $\min f(x^j)$, $j = 1, \ldots, i$.
i.e., value of best point seen to date.

Have lower bound: min value of piecewise $\underbrace{\text{linear underestimates}}_{}$ of f.

~~It can be very~~

This is a "bundle of information" provided by earlier iterates.

It can be very slow to move directly to the ~~soln~~ minimizer of the underestimator.

So use "proximal" term, $\|x - x^i\|^2$. ~~~~

So:
$$\min \quad z + \tfrac{1}{2}\mu \|x - x^i\|^2 \qquad \text{(QP)}$$
$$\text{s.t.} \quad z \geq \text{underestimator function, } \check{f}(x).$$

Then we don't move too far.

Let $\bar{x}$ be the point that solves this quadratic problem (QP).

If $f(\bar{x})$ is too large compared to z (so $\check{f}$ is not a good approximation to f),

REJECT the step, set $x^{i+1} = x^i$, update $\check{f}$ with a constraint from $\bar{x}$ and repeat

Else,

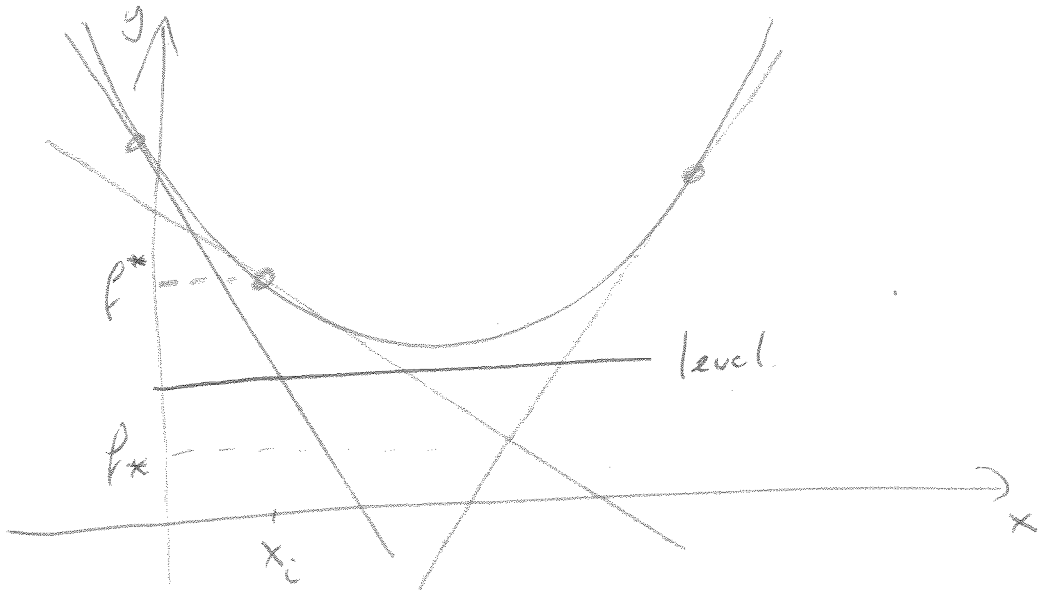
Set x^{i+1} ~~to be~~ $= \bar{x}$, update $\check{f}$ with a constraint from $\bar{x}$, and repeat.

Repeat until upper and lower bounds are close enough.

Difficulty: hard to choose μ. Should probably be dynamically updated. Nonetheless, algorithm works well.

References:

Proximal level bundle methods:



Optimal value is between f_* and f^*.

Look for proximal point with z-value $\leq \lambda f_* + (1-\lambda) f^*$ in the model, $\check{f}$.

So:
$$\min \ \tfrac{1}{2} \| x - x^i \|^2$$
$$\text{s.t.} \quad z \leq \lambda f_* + (1-\lambda) f^*$$
$$z \geq \check{f} \quad \text{underestimator of } f\text{, as a function in } x.$$

i.e. $z \geq f(x^j) + \xi^T (x - x^j)$ [illegible]

Typically, get convergence in $O\left(\frac{K}{\varepsilon}\right)$ iterations, where ε = desired accuracy, K = diameter of space = $\max\{\|x-y\| : x, y \in S\}$
Work well in practice.