· Methods for non convex problems

convex optimization methods are (roughly) always global, always fast

for general non-convex problems, we have to give up one

local optimization methods: fast. may not find global optimal, even they do, cannot certify the optimality

global optimization methods: find global solution (and certify it) but are not always fast (often slow)

· Sequential convex programming

a local optimization method for nonconvex problem that leverages convex optimization

(convex parts of a problem are handled "exactly" and efficiently)

SCP is a heuristic

it can fail to find optimal or even feasible point

result often depend on starting point

SCP often works well i.e. find a feasible point with good, if not optimal, object value

· Non convex problem

$$
\begin{aligned}
\min.\ & f_0(x) \\
\text{s.t.}\ & f_i(x) \le 0 \\
& h_i(x) = 0
\end{aligned}
$$

f_i: possibly non convex

h_i: possibly non affine

· Basic idea of SCP

maintain estimate of solution $x^{(k)}$, and convex trust region $T^{(k)}$

form convex approximation $\tilde{f}_i$ of f_i over trust region $T^{(k)}$

form affine approximation $\tilde{h}_i$ of h_i over trust region $T^{(k)}$

$x^{(k+1)}$ is the optimal point of the approximate convex problem

$$
\begin{aligned}
\min.\ & \tilde{f}_0(x) \\
\text{s.t.}\ & \tilde{f}_i(x) \le 0 \\
& \tilde{h}_i(x) = 0 \\
& x \in T^{(k)}
\end{aligned}
$$

(make sure that the convex / affine approximation are still valid)

- Trust region

typical trust region is box around current point

$$\mathcal{T}^{(k)} = \{ x \mid |x_i - x_i^{(k)}| \le \rho_i \}$$

(if x_i appears only in convex inequalities and affine equalities, can take $\rho_i = \infty$)

- Affine and convex approximation via Taylor expansion

first order taylor expansion:

$$\hat{f}(x) = f(x^{(k)}) + \nabla f(x^{(k)})^T (x - x^{(k)})$$

second order taylor expansion:

$$\hat{f}(x) = f(x^{(k)}) + \nabla f(x^{(k)})^T (x - x^{(k)}) + \frac{1}{2}(x - x^{(k)})^T P (x - x^{(k)})$$

$$P = \left(\nabla^2 f(x^{(k)}) \right)_+ \quad \text{(psd part of hessian)}$$

taylor expansion gives local approximations, which don't depend on trust region
(near $x^{(k)}$)

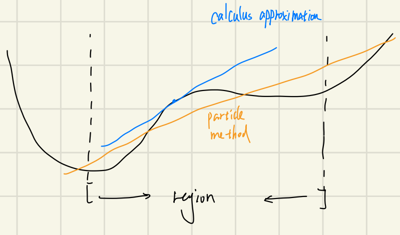
- particle method

choose points $z_1, \dots, z_p \in \mathcal{T}^{(k)}$ (eg. all vertices, some vertices, grid, random)

evaluate $y_i = f(z_i)$

fit data (z_i, y_i) with affine / convex functions (using convex optization)

(regional approximation)



- Fitting functions

fit quadratic function to data (z_i, y_i)

$$\min \sum_i \left((z_i - x^{(k)})^T P (z_i - x^{(k)}) + q^T (z_i - x^{(k)}) + r - y_i \right)^2$$

$$\text{s.t.} \quad P \succeq 0$$

can use other objectives / add convex constraints

this problem is solved at each SCP step for each non convex constraint

- Quasi-linearization

a cheap and simple method for affine approximation

write $h(x) = A(x)\,x + b(x)$. use $\hat{h}(x) = A(x^{(k)})\,x + b(x^{(k)})$

eg. $h(x) = \frac{1}{2} x^T P x + q^T x + r = (\frac{1}{2} Px + q)^T x + r$

Eg. non convex QP

$$\min.\ f(x) = \frac{1}{2} x^T P x + q^T x \qquad P \text{ is symmetric but not psd}$$
$$\text{s.t.}\ \ \|x\|_\infty \le 1 \qquad (x_i^2 \le 1)$$

use approximation

$$f(x^{(k)}) + (Px^{(k)} + q)^T (x - x^{(k)}) + \frac{1}{2} (x - x^{(k)}) P_+ (x - x^{(k)})$$

lower bound via Lagrange dual

$$g(\lambda, \nu) = \inf_x \mathcal{L}(x, \lambda, \nu)$$
$$= \inf_x f_0(x) + \textstyle\sum \lambda_i f_i(x) + \sum \nu_i h_i(x) \quad \text{(inf of affine function of } \lambda, \nu)$$

$$\mathcal{L}(x, \lambda) = \frac{1}{2} x^T P x + q^T x + \sum_{i=1}^{n} \lambda_i (x_i^2 - 1)$$
$$= \frac{1}{2} x^T (P + 2\,\mathrm{diag}(\lambda))\, x + q^T x - \mathbf{1}^T \lambda$$

$$g(\lambda) = \inf_x \mathcal{L}(x, \lambda) = \begin{cases} -\inf & \text{if } P + 2\,\mathrm{diag}(\lambda) \text{ has a negative eigenvalue} \\ & \end{cases}$$

$$\nabla_x \mathcal{L} = (P + 2\,\mathrm{diag}(\lambda))\, x + q := 0$$
$$x = -[P + 2\,\mathrm{diag}(\lambda)]^{-1} q$$
$$g(\lambda) = -\frac{1}{2} q^T (P + 2\,\mathrm{diag}(\lambda))^{-1} q - \mathbf{1}^T \lambda$$

the dual problem

$$\max_{\lambda}.\ -\frac{1}{2} q^T (P + 2\,\mathrm{diag}(\lambda))^{-1} q - \mathbf{1}^T \lambda$$
$$\text{s.t.}\ \ \lambda \succeq 0$$
$$P + 2\,\mathrm{diag}(\lambda) \succ 0$$

- Some issues

approximate convex problem may be infeasible

how to evaluate progress when $x^{(k)}$ not feasible

($f_0(x^{(k)})$ $f_i(x^{(k)})_+$ $|h_i(x^{(k)})|$)

controlling the trust region size

ρ too large: approximation are poor. leading to bad choice of $x^{(k+1)}$

ρ too small: approximations are good, but progress are slow

easily trapped in local minimum

- Exact Penalty formulation

Instead of original problem, solve the unconstrained problem

$$\min.\ \phi(x) = f_0(x) + \lambda\left(\sum_i f_i(x)_+ + \sum_i |h_i(x)|\right)$$ not squared !!!

A barrier method forces you to stay in the interior

A penalty function allows you to wander outside the feasible set, but will charge you

for SCP, use convex approximation

$$\hat{\phi}(x) = \hat{f}_0(x) + \lambda\left[\sum_i \hat{f}_i(x)_+ + \sum_i |\hat{h}_i(x)|\right]$$

when λ is big enough all $\hat{f}_i(x) \le 0$, all $\hat{h}_i(x) = 0$

when λ is not big enough, most of $\hat{h}_i(x)$ will be zero l_1 norm $\to$ sparsity

most of $\hat{f}_i(x)$ will be ≤ 0 relu penalty

- Trust region update

judge algorithm progress by decrease in ϕ, using solution $\hat{x}$ of approximate problem

decrease with approximate objective: $\hat{\delta} = \hat{\phi}(x^{(k)}) - \hat{\phi}(\hat{x})$ predicted decrease with local convex model

decrease with exact objective: $\delta = \phi(x^{(k)}) - \phi(\hat{x})$

if $\delta \approx \hat{\delta}$, the local convex model is very accurate, the trust region is too small

if $\delta \ge \alpha\hat{\delta}$, $\rho^{(k+1)} = \beta^{succ}\rho^{(k)}$, $x^{(k+1)} = \hat{x}$ ($\alpha \in (0,1)$, $\beta^{succ} \ge 1$; typical value $\alpha = 0.1$ $\beta^{succ} = 1.1$)

if $\delta < \alpha\hat{\delta}$, $\rho^{(k+1)} = \beta^{fail}\rho^{(k)}$, $x^{(k+1)} = x^{(k)}$ ($\beta^{fail} < 1$; typical value $\beta^{fail} = 0.5$)

- "Difference of Convex" programming

express problem as

$$\min.\ f_0(x) - g_0(x)$$
$$\text{s.t.}\ f_i(x) - g_i(x) \le 0$$

$\}$ f_i and g_i are convex

- Convex-concave procedure

obvious convexification at $x^{(k)}$: replace $f(x) - g(x)$ with

$$\hat{f}(x) = f(x) - g(x^{(k)}) - \nabla g(x^{(k)})^T (x - x^{(k)})$$

$f(x)$ is convex, $-g(x)$ is concave

Since $\hat{f}(x) \ge f(x)$ at all points (global upper bound), no trust region is need

- true objective at $\hat{x}$ is always better than convexified objective $\hat{f}(\hat{x})$
- true feasible set contains feasible set for convexified problems

eg. (cvx_book §7.1)

sample $x_1 \cdots x_m \in \mathbb{R}^n$ from $\mathcal{N}(0, \Sigma^{\text{true}})$

$$N(u, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-u)^T \Sigma^{-1} (x-u)\right)$$

$$-\log\, l(\Sigma) = \log\det\Sigma + \sum (x-u)^T \Sigma^{-1} (x-u)$$

$$= \underbrace{\log\det\Sigma}_{\text{concave in } \Sigma} + \underbrace{\operatorname{tr}(X\Sigma^{-1})}_{\text{convex in } \Sigma}$$

$$X = [\,|\,|\,|\,]\,[\,\equiv\,] \cdot \frac{1}{m}$$

$\Downarrow$ chang of variable

$$\log\det P^{-1} + \operatorname{tr}(XP) \quad (P = \Sigma^{-1})$$

with some positive correlation

$$\min.\ f(\Sigma) = \log\det(\Sigma) + \operatorname{tr}(\Sigma^{-1} Y)$$
$$\text{s.t.}\ \Sigma_{ij} \ge 0$$

linearize $\log\det(\Sigma)$

$$\hat{f}(\Sigma^{(k)}) = \log\det(\Sigma^{(k)}) + \operatorname{tr}\left((\Sigma^{(k)})^{-1}(\Sigma - \Sigma^{(k)})\right) + \operatorname{tr}(\Sigma^{-1} Y)$$