Convexity preserving operations

$\theta x_1 + (1-\theta) x_2$

* Intersection $\rightarrow C \cap D$
* Affine transform

$$AC+b = \{Ax+b \mid x \in C\}$$

* Translation $(C+b)$
* Scaling αC * perspective transform
* Nonnegative weighted sum $\sum_{i}^{k} w_i t_i$ is convex
* composition with affine mapping $f(Ax+b)$ is convex
* $g(f(x)) \rightarrow$ convex

for f convex, g convex & non-decreasing
f concave, g convex & non-increasing

* $f(x) = \max_i f_i(x)$ is convex
$g(x) = \inf_{y \in C} f(x,y)$ is convex

Strong convexity: $\nabla^2 f(x) \succeq mI$
$M \geq \|\nabla^2 f(x)\|$
$\kappa \leq \text{cond}(\nabla^2 f(x))$
$= M/m$

$\partial f(x) = \{g : f(y) \geq f(x) + (y-x)^T g\}$

Convex Function

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y)$$

Epigraph of a function.

$$\text{epi}(f) = \{(x,t) \mid x \in \text{dom}(f),\ t \geq f(x)\}$$

f is convex iff set $\text{epi}(f)$ is convex

Sub level sets

$$f_\alpha = \{x \mid f(x) \leq \alpha\}$$

Cones: $x \in C \Rightarrow ax \in C \ \forall a \geq 0$

Second order cone: $C_2 = \{(x,t)\} \ \|x\| \leq t\}$

Semidefinite cone: $S_+ = \bigcap_{x \in \mathbb{R}^N} \{A \mid x^T A x \geq 0\}$

Lower: $f(y) \geq f(x) + (y-x)^T \nabla f(x) + \frac{m}{2}\|x-y\|^2$

Upper: $f(y) \leq f(x) + (y-x)^T \nabla f(x) + \frac{M}{2}\|x-y\|^2$

$\|\nabla f(x) - \nabla f(y)\| \leq M \|x-y\|$

monotonicity: $\langle x-y, \nabla f(x) - \nabla f(y)\rangle \geq 0$

Subgradient of any convex function is monotone

Convexity condition

0^{th} $g(t) = f(x+tv)$ convex function of t for all x, v

$1^{st} \Rightarrow f(y) \geq f(x) + \nabla f(x)^T (y-x)$

$2^{nd} \Rightarrow \nabla^2 f(y) \succeq 0$ for all x

Some definition

proper: epi is non-empty
closed: epi is closed
LSC: proper + closed

Quasiconvex $f(\theta x + (1-\theta)y) \leq \max\{f(x), f(y)\}$

$S_\alpha = \{x \mid f(x) \leq \alpha\}$

Composition Rule (sum) DOESN'T HOLD

Conjugate function

$$f^*(y) = \max_x y^T x - f(x)$$

Duality $\quad \min f(x) \quad Ax-b=0 \longrightarrow \min_x \max_\lambda f(x) + \langle \lambda, Ax-b\rangle$

$\Downarrow$
$d(\lambda, \nu) = \min_x L(x, \lambda, \nu)$
$\Downarrow$ concave

Slater conditions $f(x) < \infty$, $g(x) < 0$ $h(x) = 0 \Rightarrow$ Strictly feasible

convex + slaters condition = strong duality.

Optimality conditions: KKT System.

$L(x,\lambda,\nu) = f(x) + \langle \lambda, h(x)\rangle + \langle \nu, g(x)\rangle$

$\nabla f(x^*) + \lambda \nabla h(x^*) + \nu \nabla g(x^*) = 0$ primal/dual optimality

$g(x^*) \leq 0$, $h(x^*) = 0$ } primal feasibility

$\nu \geq 0$ Dual feasibility

$\nu_i g_i(x^*) = 0$ comp slackness

FBS $\rightarrow$ $f \rightarrow \hat{x} = x^k - \tau \nabla f(x^k)$
$g \rightarrow x^{k+1} = \text{prox}_g(\hat{x}, \tau)$

Gaussian Elim $\Rightarrow O(K)$ LU/Cholesky $\rightarrow O(\sqrt{K}\epsilon) \rightarrow QR \rightarrow SVD$

Gradient Methods

$x^{k+1} = x^k - \tau \nabla f(x^k)$

$\tau < 2/M$

Backtracking / Armijo

Exact line search

$\tau = \min_\tau f(x^k + \tau d)$

Armijo Condition:

$f(x^k + \tau d) \leq f(x^k) + \alpha (\tau d)^T \nabla f(x^k)$, $\alpha < 1$

Wolfe condition $\rightarrow$ Armijo condition + curvature condition

$d^T \nabla f(x^k + \tau d) \geq \beta d^T \nabla f(x^k)$ $\alpha < \beta < 1$

ADAPTIVE stepsize $\Rightarrow$ BB $\tau^k = \dfrac{\Delta x^T \Delta x}{\Delta x^T \Delta g}$ $\xRightarrow{\text{weakly convex}}$ Nesterov $\tau < \dfrac{1}{L(\nabla f)}$ $\xRightarrow{\text{Poor conditioned}}$ Newton's Method $x^{k+1} = x^k - H^{-1} \nabla f(x^k)$ $\xRightarrow{\text{Approx Hessian}}$ Quasi-Newton

$y^{k+1} = x^{k+1} + \dfrac{\alpha^k - 1}{\alpha^{k+1}}(x^{k+1} - x^k)$

$x^{k+1} = y^k - \tau \nabla f(y^k)$

$\alpha^{k+1} = \dfrac{1 + \sqrt{4(\alpha^k)^2 + 1}}{2}$

MM $\rightarrow x^{k+1} = \arg\min_x f(x) + \langle \lambda^k, Ax+b\rangle + \frac{\tau}{2}\|Ax+b\|^2$

$\lambda^{k+1} = \lambda^k + \tau(Ax^{k+1} + b)$

Proximal Operator: $\text{prox}_f(z,\tau) = \arg\min_x f(x) + \frac{1}{2\tau}\|x - z\|^2$

Backward gradient: $x = z - \tau \partial f(x) \Rightarrow (\tau \partial f + I)^{-1} z = \text{prox}_f(z, \tau)$

Ex :- any proper convex function & some non convex

Shrink $(z,\tau) = \begin{cases} x^* = z - \tau & \text{if } x^* > 0 \\ x^* = z + \tau & \text{if } x^* < 0 \\ 0, & \text{otherwise} \end{cases}$

Scaled Lag.

$L_\tau(x,y,\lambda) = f(x) + g(y) + \frac{\tau}{2}\|Ax + By + c + \frac{1}{\tau}\lambda\|^2 \rightarrow \lambda \xrightarrow{\text{G.D.}} \hat{\lambda}$ P.P

PDHG $\min \max f(x) + \langle Ax, y\rangle - g(y)$

$\hat{x} = x_k - \tau A^T y$

$x_{k+1} = \arg\min f(x) + \frac{1}{2\tau}\|x - \hat{x}\|^2$

Predict $\bar{x} = x_{k+1} + x_{k+1} - x_k$

$y_{k+1} = \arg\min g(y) + \frac{1}{2\sigma}\|y - \hat{y}\|^2$ $\hat{y} = y_k + \tau A \bar{x}$

RANDOM STUFF $\quad \max p(x) \quad \max p(x)$ MC $\rightarrow$ Irreducible, Aperiodic

Gradient/Splitting e^{-ck} SGD $1/K$ MCMC

SGD $\rightarrow$

Strongly convex problems $\rightarrow \tau_K = \dfrac{a}{b+k}$

Weakly convex problems $\rightarrow \tau_K = \dfrac{a}{\sqrt{K+b}}$

Ergodic Averaging $x^{k+1} = \dfrac{1}{k+1}\sum x_i$

$\boxed{\text{Second order Cone} \quad f^T x \quad \|A_i x + b_i\| \leq c_i^T x + d_i}$

compute without storage / some memory

$x^{k+1} = \dfrac{k}{k+1} x^k + \dfrac{1}{k+1} x^{k+1}$

$\boxed{O(N^6 L) \quad O(N^{3.5} L)}$

gradient must be bounded

MH algorithm = Metropolis Hastings

Start with x^0

For $k = 1, 2, 3, \ldots$

Choose candidate y from $q(y \mid x^k)$

Compute acceptance probability

$\alpha = \min\left\{1, \dfrac{P(y)\, q(x^k \mid y)}{P(x^k)\, q(y \mid x^k)}\right\}$

$x^{k+1} = y$ with prob. α

otherwise $x^{k+1} = x^k$

Simulated Annealing :- $p^{1/T_k}(x^k)$

$T_k = \dfrac{1}{C \log(k + T_0)}$

$\left[\begin{array}{c}\text{Standard Form LP} \\ c^T x, \ Ax = b \ x \geq 0\end{array}\right]$ $\left[\begin{array}{c}\text{General Form} \\ \min c^T x \\ \text{subj } Gx \leq h \\ Ax = b\end{array}\right]$

Two approach

Phase 1: min t, $Ax = b$, $x \geq -t$, $t \geq 0$

Phase 2: $\max c^T x$ sub $Ax = b$, $x \geq 0$

Primal $c^T x$, $Ax = b$, $x \geq 0$

dual $\max -b^T \lambda$, $A^T \lambda \geq -c$

Adding constraint (row) to primal is like adding a variable (column) to the dual

Barrier IP $\rightarrow f_0(x) - \mu \log(x)$, $Ax = b$, $\mu \leftarrow \mu/\ldots$

Primal-dual $\rightarrow$ non linear to solve KKT system $\rightarrow$ no separate centering step

comp slackness $\nu_i x_i = \mu$

classical $\rightarrow J(x)\Delta x = -F(x)$
$x \leftarrow x + \alpha \Delta x$

Composite $\rightarrow$ predictor corrector

$J(x)\Delta x_p = -F(x)$ $J(x)\Delta x_c = -F(x + \Delta x_p)$ $(\lambda, \nu) \leftarrow (\lambda, \nu) + \alpha_d(\Delta\lambda, \Delta\nu)$

$x \leftarrow x + \alpha(\Delta x_p + \Delta x_c)$

$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ V & 0 & X \end{pmatrix}$