Subgradient method for Constrained Problems

11:31 AM

[Contents of the page:]

* Projected subgradient method: [Projected Subgradient Method] [Contents of the page:]

$$\begin{pmatrix} \min f(x) \\ \text{s.t. } x \in C \end{pmatrix}$$

$\{\nabla\}$ $\{\square\}$

$$x^{(k+1)} = \Pi\left(x^{(k)} - \alpha_k g^{(k)}\right), \quad g^{(k)} \in \partial f(x^{(k)})$$

Π: projection on C

Convergence Results: converges to neighborhood of optimal for constant step size

- diminishing nonsummable step sizes converges

Linear equality constraints:

$$\begin{pmatrix} \min f(x) \\ \text{s.t. } Ax = b \end{pmatrix} \quad C = \{x \mid Ax = b\}: \quad A\{\text{fat, full rank}\}$$

$$x^{(k+1)} = \Pi\left(x^{(k)} - \alpha_k g^{(k)}\right)$$

$\boxed{\text{Calafiore page 473:}}$

$\forall_{x \in \mathcal{X}} \forall_{z \in \mathcal{X}} \; f_0(z) \geq f_0(x) + g_x^T(z - x)$ # assumption $\mathcal{X} \subset \text{int dom} f$

g_x {subgradient}

$z = x^*$

$\forall_{x \in \mathcal{X}} \; f_0(x^*) \geq f_0(x) + g_x^T(x^* - x)$

$\Leftrightarrow \; -g_x^T(x^* - x) = g_x^T(x - x^*) \geq \underbrace{f_0(x) - f_0(x^*)}_{\text{because of optimality of } x^*} \geq 0$

$\forall_{x \in \mathcal{X}} \; g_x^T(x - x^*) \geq f_0(x) - f_0(x^*) \geq 0$ (12.73)

$\mathcal{H}_{++} = \{z : g_x^T(z - x) > 0\}$, $\mathcal{H}_{-} = \{z : g_x^T(z - x) \leq 0\}$

$\forall_{z \in \mathcal{H}_{++}} \; f_0(z) \geq f_0(x) + g_x^T(z - x) > f_0(x)$

$\rightarrow \forall_{z \in \mathcal{H}_{++}} \; f_0(z) > f_0(x)$: We do not need to look in $\mathcal{H}_{++}$

Similarly

$\forall_{z \in \mathcal{H}_{-}} \; f_0(z) \leq f_0(x)$: need to look only in $\mathcal{H}_{-}$

* Subgradient algorithm for $\begin{pmatrix} \min f(x) \\ \text{s.t. } x \in \mathcal{X} \end{pmatrix}$ → simple closed convex set; easy to take Eucledian projection

$\forall_{k \in \{0, 1, \dots\}}$

$x_{k+1} = [x_k - s_k g_k]_{\mathcal{X}}$

g_k → subgradient of f_0 at x_k

$[\cdot]_{\mathcal{X}}$:= Eucledian projection onto $\mathcal{X}$

$x_0 \in \mathcal{X}$

s_k: suitable stepsize

Proposition 12.1;

$\{\exists_{x^*} x^*\{\min\}, \forall_{x \in \mathcal{X}} \forall_{g \in \partial f_0(x)} \exists_{G \in (-\infty, \infty)} \|g\|_2 \leq G, \; \exists_R \|x_0 - x^*\| \leq R\} \Rightarrow f_{0,k}^* - p^* \leq \dfrac{R^2 + G^2 \sum_{i=0}^{k} s_i^2}{2 \sum_{i=0}^{k} s_i}$

$f_{0,k}^* = \min_{i \in \{0, \dots, k\}} \{f_0(x_i)\}$

// Clearly for square summable

// but not summable sequence

// $s_k = r/(k+1)$, $r > 0$, $\lim_{k \to \infty} f_{0,k}^* - p^* = 0$

Proof: $z_{k+1} = x_k - s_k g_k$ // update in the direction of the negative // subgradient before taking projection on // $\mathcal{X}$ is made

$\|z_{k+1} - x^*\|_2^2$

$= \|x_k - s_k g_k - x^*\|_2^2 = \left((x_k - x^*) - s_k g_k\right)^T \left((x_k - x^*) - s_k g_k\right) = \|x_k - x^*\|_2^2 + s_k^2 \|g_k\|_2^2 - 2 s_k g_k^T (x_k - x^*)$

$= \|x_k - x^*\|_2^2 + s_k^2 \|g_k\|_2^2 + 2 s_k g_k^T (x^* - x_k)$

$\left\{ g_k^T(x - x^*) \geq f_0(x) - f_0(x^*) \text{ (from 12.73)} \right.$

$\left. \Leftrightarrow f_0(x^*) - f_0(x) \geq g_k^T(x^* - x) \right\}$

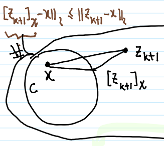
$\leq \|x_k - x^*\|_2^2 + s_k^2 \|g_k\|_2^2 + 2 s_k \left(\underbrace{f_0(x^*)}_{} - f_0(x_k)\right)$

// $p^* = f_0(x^*)$; $x^*\{\min\}$

$x_{k+1} = [z_{k+1}]_{\mathcal{X}}$

[illegible], $\|x_{k+1} - x^*\|_2$ [illegible] change [illegible]. At first see:

$\forall_{x \in \mathcal{X}} \|x_{k+1} - x\|_2 = \|[z_{k+1}]_{\mathcal{X}} - x\|_2 \leq \|z_{k+1} - x\|_2$



as $[z_{k+1}]_{\mathcal{X}} = \arg\min_{y \in \mathcal{X}} \frac{1}{2}\|y - z_{k+1}\|_2^2$, so

$\|z_{k+1} - x\|_2 \geq \|[z_{k+1}]_{\mathcal{X}} - x\|_2$

$x^* = x \Rightarrow$

$\|x_{k+1} - x^*\|_2^2 \leq \|z_{k+1} - x^*\|_2^2 \leq \|x_k - x^*\|_2^2 + s_k^2 \|g_k\|_2^2 + 2 s_k (p^* - f_0(x_k))$

Note: after this stage the convergence proof is exactly similar to Basic Subgradient method in [Convergence Proof: Subgradient Method]

$\rightarrow \|x_{k+1} - x^*\|_2^2 \leq \|x_k - x^*\|_2^2 + s_k^2 \|g_k\|_2^2 - 2 s_k (f_0(x_k) - p^*)$ ⊖

$x^* = x \Rightarrow$

$$\|x_{k+1}-x^*\|_2^2 \leqslant \|z_{k+1}-x^*\|_2^2 \leqslant \|x_k-x^*\|_2^2 + s_k^2\|g_k\|_2^2 + 2s_k(p^* - f_0(x_k))$$

Note: after this stage the convergence proof is exactly similar to Basic Subgradient method in [Convergence Proof: Subgradient Method]

$$\rightarrow \|x_{k+1}-x^*\|_2^2 \leqslant \|x_k-x^*\|_2^2 + s_k^2\|g_k\|_2^2 - 2s_k(f_0(x_k)-p^*) \quad ①$$

Similarly:

$$// \; k := k-1 \Rightarrow \|x_k - x^*\|_2^2 \leqslant \|x_{k-1}-x^*\|_2^2 + s_{k-1}^2\|g_{k-1}\|_2^2 - 2s_{k-1}(f_0(x_{k-1})-p^*)$$

$\vdots$

$$// \; k := 0 \Rightarrow \|x_1 - x^*\|_2^2 \leqslant \|x_0-x^*\|_2^2 + s_0^2\|g_0\|_2^2 - 2s_0(f_0(x_0)-p^*)$$

②

$$\|x_{k+1}-x^*\|_2^2 \leqslant \|x_0-x^*\|_2^2 + \sum_{i=0}^{k} s_i^2\|g_i\|_2^2 - 2\sum_{i=0}^{k} s_i(f_0(x_i)-p^*)$$

$$\Leftrightarrow 2\sum_{i=0}^{k} s_i(f_0(x_i)-p^*) \leqslant \|x_0-x^*\|_2^2 - \underbrace{\|x_{k+1}-x^*\|_2^2}_{\in \mathbb{R}_- \text{ so বাদ দিলে বড় হবে}} + \sum_{i=0}^{k} s_i^2\|g_i\|_2^2$$

$$\leqslant \|x_0-x^*\|_2^2 + \sum_{i=0}^{k} s_i^2\|g_i\|_2^2$$

$$\leqslant R^2 + \sum_{i=0}^{k} s_i^2\|g_i\|_2^2 \quad \{\because \|x_0-x^*\| \leqslant R : \text{given}\}$$

$$\{\text{now}, \forall_{i\in\{0,\dots,k\}} \quad f_{0,k}^* = \min_{i=1,\dots,k} f_0(x^{(i)}) \leqslant f(x^{(i)})$$

$$\rightarrow \forall_{i\in\{0,\dots,k\}} \quad f_{0,k}^* - p^* \leqslant f_0(x^{(i)}) - p^*$$

$$\rightarrow \forall_{i\in\{0,\dots,k\}} \quad s_i(f_{0,k}^* - p^*) \leqslant s_i(f_0(x^{(i)}) - p^*) \quad \{\because s_i > 0\}$$

$$\rightarrow \sum_{i=k}^{0} s_i(f_{0,k}^* - p^*) \leqslant \sum_{i=k}^{0} s_i(f_0(x^{(i)}) - p^*)$$

$$\Rightarrow 2\sum_{i=0}^{k} s_i(f_{0,k}^* - p^*) \leqslant 2\sum_{i=0}^{k} s_i(f_0(x_i)-p^*) \leqslant R^2 + \sum_{i=0}^{k} s_i^2\|g_i\|_2^2$$

এখানে index free

$$\Rightarrow 2(f_{0,k}^* - p^*)\Big(\sum_{i=0}^{k} s_i\Big) \leqslant R^2 + \sum_{i=0}^{k} s_i^2\|g_i\|_2^2 \leqslant R^2 + \sum_{i=0}^{k} s_i^2 G^2 \quad \{\because \forall_i \; \|g_i\|_2 \leqslant G\}$$

$$\rightarrow (f_{0,k}^* - p^*) \leqslant \left(\frac{R^2 + G^2\sum_{i=0}^{k} s_i^2}{2\sum_{i=0}^{k} s_i}\right)$$

The optimality bound for different step sizes can be found at: [Optimality bound for different step sizes]

Clearly $s_i = \frac{\gamma}{i+1}$, $\gamma \in \mathbb{R}_{++}$, $\lim_{k\to\infty} \sum_{i=0}^{k} s_i^2 = 0$, $\lim_{k\to\infty}\sum_{i=0}^{k} s_i = \infty$

[Alternate Subgradient Method] [Contents of the page:]

The alternate subgradient Method

$$p^* = \Big(\forall f_0(x) \;\; \wedge \; \forall_{i\in\{1,\dots,m\}} f_i(x) \leqslant 0\Big) \quad \{\in D; \text{ may be } \mathbb{R}^n\}$$

$$// \; h(x) = \max_{i\in\{1,\dots,m\}} f_i(x) \leqslant 0$$

$$p^* = \Big(\forall_{x\in\mathbb{R}^n} f_0(x) \wedge h(x) \leqslant 0\Big)$$

Algorithm:

$$\forall_{k\in\{0,1,\dots\}} \; x_{k+1} = x_k - s_k g_k, \quad s_k \in \mathbb{R}_{++}$$

$$g_k \in \begin{cases} \partial f_0(x_k), & \text{if } h(x_k) \leqslant 0 \quad \text{// normal subgradient alg for unconstrained optimization} \\ \partial h(x_k), & \text{if } h(x_k) > 0 \end{cases}$$

% Constraint set এর বাইরে current iterate x_k চলে গেলে
% $\partial h(x_k)$ দিয়ে (·) ভিতরে ঢোকার চেষ্টা, $h(x_k)>0$ হলে $h(x_{k+1}) \leqslant 0$ এ ঢোকাতে হলে আমাদেরকে $h(\cdot)$
% যেদিকে কমে, সেদিকে যেতে হবে, i.e. $-g_k$ বরাবর যেতে হবে if $g_k = \nabla_x h(x_k)$, now
% subgradient, gradient এর generalization বলে intuitively এই logic follow করা হচ্ছে

// $h(x_k) = \max_{i\in\{1,\dots,m\}} f_i(x_k)$ এর max rule subgradient calculus ব্যবহার
// করে one subgradient বের করা যায়.

$$f_{0,k}^* = \min\{f_0(x_i) : \forall_{i\in\{0,\dots,k\}} \; x_i \text{ feasible}\}, \quad h(x_i) \leqslant 0$$

x_{sf} : a strictly feasible point $\Leftrightarrow h(x_{sf}) < 0$

Convergance of alternate subgradient :

$$\Big\{\exists_{x_{sf}} h(x_{sf}) < 0, \; \exists_{x^*\{\forall\}} f_0(x^*) \in (-\infty,\infty), \; \exists_{R\{\text{finite}\}}\big(\|x_0-x^*\|_2 \leqslant R, \; \|x_0 - x_{sf}\|_2 \leqslant R\big), \; \forall_k \exists_{G\{\text{finite}\}} \|g_k\|_2 \leqslant G\Big\} \Rightarrow \lim_{k\to\infty}(f_{0,k}^* = p^*)$$

Proof: By contradiction. let Given $\wedge$ ¬Goal

$$\Big(\lim_{k\to\infty} f_{0,k}^* > p^*\Big) \quad \{\text{by defn. } p^* \text{ এর চাইতে কম কিছু হতে পারে না}\}$$

$$\equiv \forall_k \; f_{0,k}^* > p^*$$

$$\equiv \exists_{\varepsilon>0}\forall_k \; f_{0,k}^* \geqslant p^* + \varepsilon \quad // \; f_{0,k}^* \geqslant p^* + \varepsilon_k \equiv f_{0,k}^* - \varepsilon_k \geqslant p^* \;\therefore \forall_k \; f_{0,k}^* - \varepsilon_k \geqslant p^* \Leftrightarrow \min_k\{f_{0,k}^*\} + \underbrace{\min_k\{-\varepsilon_k\}}_{= -\max_k\{\varepsilon_k\}} \geqslant p^* \equiv \min_k\{f_{0,k}^*\} - \underbrace{\max_k\{\varepsilon_k\}}_{\varepsilon} \geqslant p^*$$

$$\equiv \exists_{\varepsilon>0}\forall_k \; \min\{f_0(x_i) : x_i \text{ feasible}, i=0,\dots,k\} \geqslant p^* + \varepsilon \qquad // \equiv \min_k\{f_{0,k}^*\} \geqslant p^* + \varepsilon$$

$$\equiv \exists_{\varepsilon>0}\forall_k \forall_{i=0,\dots,k} \forall_{x_i:\text{feasible}} \; f_0(x_i) \geqslant p^* + \varepsilon \quad (\text{eq:per\_absurdum\_statement}) \qquad // \equiv \forall_k \; f_{0,k}^* \geqslant p^* + \varepsilon$$

// so : $\boxed{\varepsilon = \max_k\{\varepsilon_k\}}$

$$\equiv \text{Given} \wedge \exists_{\varepsilon>0}\forall_k\forall_{i=0,\dots,k}\forall_{x_i:\text{feasible}} \; f_0(x_i) \geqslant p^* + \varepsilon$$

$$\text{let } \forall_{\theta\in(0,1]} \; \tilde{x} = (1-\theta)x^* + \theta x_{sf}$$

$$f_0\{D\} \rightarrow f_0(\tilde{x}) = f((1-\theta)x^* + \theta x_{sf}) \leqslant (1-\theta)f(x^*) + \theta f(x_{sf}) = f(x^*) + \theta\underbrace{(f(x_{sf}) - f(x^*))}_{\geqslant 0 \text{ // equality হতে পারে as } x_{sf}=x^* \text{ হতে পারে}}$$

$$\theta := \min\Big\{1, \frac{\varepsilon}{2}\cdot\frac{1}{(f_0(x_{sf})-p^*)}\Big\}$$

// note that how cleverly this is choosen
// if $x_{sf} \neq x^*$, then finite, even if $x_{sf} = x^*$ then it is
// infinite but $\min\{1,\infty\} = 1$ still
// finite, 1 এর চাইতে বড় হলেও min 1 হবে

$$f_0(\tilde{x}) \leqslant \underset{p^*}{f(x^*)} + \min\Big\{1, \frac{\varepsilon}{2}\frac{1}{(f_0(x_{sf})-p^*)}\Big\}\big(f(x_{sf}) - \underset{p^*}{f(x^*)}\big)$$

$\leqslant p^* + \min\{ f(x_{sf}) - p^*, \frac{\varepsilon}{2} \}$ // using $\underset{\{\in \mathbb{R}_+\}}{a} \min\{b,c\} = \min\{ab, ac\}$

$\rightarrow f_0(\tilde{x}) \leqslant p^* + \frac{\varepsilon}{2}$ // $\because x \leqslant \min\{a,b\} \leftrightarrow x \leqslant a \wedge x \leqslant b$

$\rightarrow f_0(\tilde{x}) - p^* \leqslant \frac{\varepsilon}{2}$.

$\therefore 0 \leqslant f_0(\tilde{x}) - p^* \leqslant \frac{\varepsilon}{2} \leftrightarrow \tilde{x}$ is $\frac{\varepsilon}{2}$ suboptimal (eq:suboptimality)

Again:

$$h(\tilde{x}) = h((1-\theta)x^* + \theta x_{sf}) \leqslant \underbrace{(1-\theta)\, h(x^*)}_{\leqslant 0\ [\because x^* \text{ হচ্ছে feasible হতে হবে}]} + \theta\, h(x_{sf}) \leqslant \theta\, h(x_{sf}) = -\mu$$

$\theta \in (0,1]$, <0, $\mu \in \mathbb{R}_{--}$, <0 // $\in \mathbb{R}_-$ হবে কারণ দিতে পারি

$\rightarrow h(\tilde{x}) \leqslant -\mu$ (eq:strict_feasibility)

$\forall_{i \in \{0,1,\dots,k\}} (h(x_i) \leqslant 0 \vee h(x_i) > 0)$

First, consider $h(x_i) \leqslant 0 \leftrightarrow x_i$ feasible so: • eq:per_absurdum_statement থেকে $f_0(x_i) \geqslant p^* + \varepsilon_k \rightarrow f_0(x_i) - p^* \geqslant \varepsilon$

• algorithm এর definition থেকে $g_i \in \partial f_0(x_i)$

(eq:suboptimality) কে -1 দ্বারা গুণ দিয়ে $-\frac{\varepsilon}{2} \leqslant -f_0(\tilde{x}) + p^* \leqslant 0$

$\rightarrow -f_0(\tilde{x}) + p^* \geqslant -\frac{\varepsilon}{2}$ $\rightarrow (+)$

$$\boxed{f_0(x_i) - f_0(\tilde{x}) \geqslant \frac{\varepsilon}{2}}$$ eq:difference_btn_f0xi_and_f0xTilde

$\because x_{i+1} = x_i - s_i g_i$, $g_i \in \partial f_0(x_i)$ [$\because x_i$ feasible]

$$\|x_{i+1} - \tilde{x}\|_2^2 = \|x_i - s_i g_i - \tilde{x}\|^2$$

$$= ((x_i - \tilde{x}) - s_i g_i)^T ((x_i - \tilde{x}) - s_i g_i)$$

$$= \|x_i - \tilde{x}\|_2^2 + s_i^2 \|g_i\|_2^2 - 2 s_i g_i^T (x_i - \tilde{x}) = \|x_i - \tilde{x}\|_2^2 + s_i^2 \|g_i\|_2^2 + 2 s_i g_i^T (\tilde{x} - x_i)$$ (eq:norm_difference)

// this is just matrix algebra
// so far, so will hold for infeasible x_i too

now: $g_i \in \partial f_0(x_i) \leftrightarrow \forall_x\ f_0(x) \geqslant f_0(x_i) + g_i^T (x - x_i)$

$\therefore x = \tilde{x}$ yields $f_0(\tilde{x}) \geqslant f_0(x_i) + g_i^T (\tilde{x} - x_i)$

$\overset{\text{or}}{\equiv} f_0(\tilde{x}) - f_0(x_i) \geqslant g_i^T (\tilde{x} - x_i)$

$$\leqslant \|x_i - \tilde{x}\|_2^2 + s_i^2 \|g_i\|_2^2 + 2 s_i (f_0(\tilde{x}) - f_0(x_i))$$

now: eq:difference_btn_f0xi_and_f0xTilde থেকে (-1)দ্বারা গুণ দিয়ে: $f_0(\tilde{x}) - f_0(x_i) \leqslant -\frac{\varepsilon}{2}$

$$\leqslant \|x_i - \tilde{x}\|_2^2 + s_i^2 \|g_i\|_2^2 - s_i \varepsilon \qquad \therefore \boxed{\|x_{i+1} - \tilde{x}\|_2^2 \leqslant \|x_i - \tilde{x}\|_2^2 + s_i^2 \|g_i\|_2^2 - s_i \varepsilon}$$

Now consider:

x_i {infeasible} $\leftrightarrow h(x_i) > 0$

eq:strict_feasibility X(-1): $-h(\tilde{x}) \geqslant \mu$

(+) $h(x_i) - h(\tilde{x}) \geqslant \mu \overset{\text{or}}{\equiv} h(\tilde{x}) - h(x_i) \leqslant -\mu$

in this case:

$$\|x_{i+1} - \tilde{x}\|_2^2 = \|x_i - \tilde{x}\|_2^2 + s_i^2 \|g_i\|_2^2 + 2 s_i g_i^T (\tilde{x} - x_i)$$ eq:norm_difference থেকে

$g_i \in \partial h(x_i) \Rightarrow h(\tilde{x}) \geqslant h(x_i) + g_i^T (\tilde{x} - x_i) \overset{\text{or}}{\equiv} h(\tilde{x}) - h(x_i) \geqslant g_i^T (\tilde{x} - x_i)$

$$\leqslant \|x_i - \tilde{x}\|_2^2 + s_i^2 \|g_i\|_2^2 + 2 s_i (h(\tilde{x}) - h(x_i))$$

$$\leqslant \|x_i - \tilde{x}\|_2^2 + s_i^2 \|g_i\|_2^2 - 2 s_i \mu$$

$$\boxed{\|x_{i+1} - \tilde{x}\|_2^2 \leqslant \|x_i - \tilde{x}\|_2^2 + s_i^2 \|g_i\|_2^2 - 2 s_i \mu}$$

so x_i feasible বা infeasible যেটাই হোক না কেন,

$$\|x_{i+1} - \tilde{x}\|_2^2 \leqslant \|x_i - \tilde{x}\|_2^2 + s_i^2 \|g_i\|_2^2 - s_i \beta$$

$\beta = \underbrace{\min\{2\mu, \varepsilon\}}$ ফিক্সড ধরেছি independent of k. by construction

// যেমন minimizing logic is different (it is an implication)
// than conjunctive minimizing logic. (which was an equivalence)
// $\mu > 0, \varepsilon > 0 \Rightarrow -2 s_i \mu < 0 \wedge -s_i \varepsilon < 0$
// $s_i > 0$
// so feasible বা infeasible যাই হোক না কেন
// $\beta = \min\{2\mu, \varepsilon\} \leqslant \varepsilon \wedge \beta = \min\{2\mu, \varepsilon\} \leqslant 2\mu$
// $-s_i \beta \geqslant -s_i \varepsilon \wedge -s_i \beta \geqslant -2 s_i \mu$ [negative $-2s_i$ দ্বারা গুণ করার inequality flipped]
// so no matter what (:) is always true.

$$\therefore \forall_{i \in \{0,\dots,k\}} \|x_{i+1} - \tilde{x}\|_2^2 \leqslant \|x_i - \tilde{x}\|_2^2 + s_i^2 \|g_i\|_2^2 - s_i \beta$$

$i = k$: $\|x_{k+1} - \tilde{x}\|_2^2 \leqslant \|x_k - \tilde{x}\|_2^2 + s_k^2 \|g_k\|_2^2 - s_k \beta$

$i = k-1$: $\|x_k - \tilde{x}\|_2^2 \leqslant \|x_{k-1} - \tilde{x}\|_2^2 + s_{k-1}^2 \|g_{k-1}\|_2^2 - s_{k-1} \beta$

$\vdots$

$i = 0$: $\|x_1 - \tilde{x}\|_2^2 \leqslant \|x_0 - \tilde{x}\|_2^2 + s_0^2 \|g_0\|_2^2 - s_0 \beta$

Backward substitution করে পাই:

$$\|x_{k+1} - \tilde{x}\|_2^2 \leqslant \|x_0 - \tilde{x}\|_2^2 + \sum_{i=0}^{k} s_i^2 \|g_i\|_2^2 - \beta \sum_{i=0}^{k} s_i$$

By assumption, $\|x_0 - \tilde{x}\|_2 \leqslant R$, $\forall_i\ \|g_i\|_2 \leqslant G$

$$\rightarrow \underbrace{0 \leqslant \|x_{k+1} - \tilde{x}\|_2^2}_{\text{Norm nonnegativity}} \leqslant R^2 + G^2 \sum_{i=0}^{k} s_i^2 - \beta \sum_{i=0}^{k} s_i$$

$$\rightarrow \beta \sum_{i=0}^{k} s_i \leqslant R^2 + G^2 \sum_{i=0}^{k} s_i^2$$

this is a contradiction as $k \to \infty$, for a square-summable, but non-summable sequence $\{s_i\}_{i=0}^{k}$, as $k \to \infty$, the LHS becomes $+\infty$, but RHS becomes R^2 which is finite.

$\therefore$ Alternate subgradient method converges!

(Proved)

[Linear inequality constraints problem using projected subgradient] [Contents of the page:]

Linear inequality constraints problem using projected subgradient:

$\begin{pmatrix} \forall\ f(x) \\ \neg\ Ax = b \end{pmatrix}$

the iteration will be: $x^{(k+1)} = \Pi_{\{\square : A\square = b\}} \left(x^{(k)} - \alpha_k g^{(k)} \right)$

Projection on Az=b থেকে পাবো:

the iteration will be: $x^{(k+1)} = \Pi_{\{\square : A\square = b\}}(x^{(k)} - \alpha_k g^{(k)})$

[Projection on Az=b] (এখানে যাও:

$$\Pi_{\{\square : A\square = b\}}(z) = \left((I - A^T(AA^T)^{-1}A)\square + A^T(AA^T)^{-1}b\right)(z)$$

$$= (I - A^T(AA^T)^{-1}A)z + A^T(AA^T)^{-1}b$$

$$x^{(k+1)} = \Pi_{\{\square : A\square = b\}}(x^{(k)} - \alpha_k g^{(k)})$$

$$= (I - A^T(AA^T)^{-1}A)(x^{(k)} - \alpha_k g^{(k)}) + A^T(AA^T)^{-1}b$$

$$= x^{(k)} - \alpha_k g^{(k)} - A^T(AA^T)^{-1}\underbrace{Ax^{(k)}}_{= b} + A^T(AA^T)^{-1}A\alpha_k g^{(k)} + A^T(AA^T)^{-1}b$$

এখানে $x^{(k)} = \Pi^{\{\square : A\square = b\}}(\square)$ so $x^{(k)} \in \{\square : A\square = b\}$

$$= x^{(k)} - \alpha_k I g^{(k)} - A^T\cancel{(AA^T)^{-1}b} + \alpha_k A^T(AA^T)^{-1}A g^{(k)} + \cancel{A^T(AA^T)^{-1}b}$$

$$= x^{(k)} - \alpha_k \underbrace{(I - A^T(AA^T)^{-1}A)g^{(k)}}_{\Pi_{\mathcal{N}(A)}(g^{(k)})}$$

Note $\Pi_{\{\square : A\square = 0\}}(z) = (I - A^T(AA^T)^{-1}Az + A^T(AA^T)^{-1}b = (I - A^T(AA^T)^{-1}A)z = \Pi_{\mathcal{N}(A)}(z)$, where $\{\square : A\square = 0\} = \mathcal{N}(A)$, $0 = b$, $b = 0$

$$= x^{(k)} - \alpha_k \Pi_{\mathcal{N}(A)}(g^{(k)})$$

Numerical Example: $\min \|x\|_1$ s.t. $Ax = b$

[subgradient of l-1 norm] (এখানে যাও: $\partial f(x) = \sum_{i=1}^{n} \left(e_i \, \mathrm{sgn}(x_i) \{x_i \neq 0\} + [-1,1] \{x_i = 0\}\right)$

so $g(x) = \mathrm{sgn}(x) = \begin{bmatrix} \mathrm{sgn}(x_1) \\ \vdots \\ \mathrm{sgn}(x_n) \end{bmatrix} \in \partial f(x)$

now the projected subgradient alg:

$$x^{(k+1)} = x^{(k)} - \alpha_k \Pi_{\mathcal{N}(A)}(g^{(k)})$$

$$= x^{(k)} - \alpha_k \Pi_{\mathcal{N}(A)}(\mathrm{sgn}(x))$$

*Projected subgradient for dual problem: [Projected Subgradient for dual problem] [Contents of the page:]

//Famous application of projected subgradient method, in generally there is no reason to solve dual instead of primal. but for specific problems, there can be advantage.

Consider:

$$\begin{pmatrix} \min f_0(x) \\ \text{s.t.} \\ \forall_{i \in \{1,\dots,m\}} f_i(x) \leq 0 \end{pmatrix} \text{ \# : convex}$$

An interpretation of this $f_i(x) \leq 0$ is if the consumption of resource i,
for within budget $f_i(x) \leq 0$.
for over budget $f_i(x) > 0$

$$L(x, \lambda) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x)$$

Assumption:

1) $\forall_{\lambda \succeq 0} \left(x^*(\lambda) = \operatorname{argmin}_x L(x,\lambda)\right) : !$ // When this might happen? say if $f_0(x) : \cup_{\text{strongly}} \Rightarrow \left(f_i(x)_{:\cup_{\text{strongly}}} + \left(\sum \lambda_i f_i(x)\right)_{:\cup}\right) : \cup_{\text{strongly}}$ as $\cup_{\text{strongly}} + \cup = \cup_{\text{strongly}}$

$\exists_{m>0} \left(f_0(x) - \frac{m}{2}\|x\|_2^2\right) : \cup$

2) Slater's condition holds, $p^* = d^*$, so we can find optimal solution to primal problem by solving the dual problem (find λ^*) and then set $x^* = x^*(\lambda^*)$

quick proof:

$f(x) : \cup_{\text{strongly}} \leftrightarrow \exists_{m>0} \left(f(x) - \frac{m}{2}\|x\|_2^2\right) : \cup$

$g(x) : \cup \leftrightarrow g(x) : \cup$

$\left(f(x) - \frac{m}{2}\|x\|_2^2 + g(x)\right) : \cup$

$\leftrightarrow \left(f(x) + g(x)\right) - \frac{m}{2}\|x\|_2^2 : \cup$

$\leftrightarrow \exists_m \; f(x) + g(x) - \frac{m}{2}\|x\|_2^2 : \cup$

$\leftrightarrow f(x) + g(x) : \cup_{\text{strongly}}$

$$g(\lambda) = \inf_x L(x,\lambda) = f_0(x^*(\lambda)) + \sum_{i=1}^{m} \lambda_i f_i(x) \quad \therefore -g(\lambda) = -\inf_x \underbrace{L(x,\lambda)}_{\text{affine in } \lambda} = -\left(-\sup(-L(x,\lambda))\right) = \underbrace{\sup_x (-L(x,\lambda))}_{\text{supremum of a family of affine functions of } \lambda \text{ indexed by } x}$$

the dual problem is:

$$\begin{pmatrix} \max g(\lambda) \\ \text{s.t. } \lambda \succeq 0 \end{pmatrix} = -\underbrace{\begin{pmatrix} \min -g(\lambda) \\ \text{s.t. } \lambda \succeq 0 \end{pmatrix}}_{\text{convex optimization problem say (problem 1)}}$$

So, by projected subgradient problem can be solved as:

$$\lambda^{(k+1)} = \Pi_{\square \succeq 0}(\lambda^{(k)} - \alpha_k h^{(k)}) = (\lambda^{(k)} - \alpha_k h^{(k)})_+ \quad \{x_+ = (\max\{x_i, 0\})_{i=1}^{n}\}$$

$h^{(k)} \in \partial(-g)(\lambda^{(k)})$; now [eq: subgradient for negative dual function] (এখানে যাও: $(-f_i(x^*(\lambda)))_{i=1}^{m} \in \partial(-g(\lambda))$

$(x^*(\lambda) = \operatorname{argmin}_x L(x,\lambda)) : !$

$$h^{(k)} = (-f_i(x^*(\lambda^{(k)})))_{i=1}^{m} \quad // \quad x^*(\lambda^{(k)}) = \operatorname{argmin}_x \left(f_0(x) + \sum_{i=1}^{m} \lambda_i^{(k)} f_i(x)\right) = x^{(k)}$$ (to avoid all the $*, \lambda^{(k)}$ etc in the notation)

$$= (-f_i(x^{(k)}))_{i=1}^{m}$$

(Our assumptions imply that $-g$ has only one element in its subdifferential, which means g is differentiable. Differentiability means that a small enough constant step size will yield convergence. In any case, the projected subgradient method can be used in cases where the dual is nondifferentiable.)

$$\lambda^{(k+1)} = (\lambda^{(k)} - \alpha_k h^{(k)})_+ = \left(\lambda^{(k)} - \alpha_k \left((-f_i(x^{(k)}))_{i=1}^{m}\right)\right)$$

$$x^{(k)} = \operatorname{argmin}_x \left(f_0(x) + \sum_{i=1}^{m} \lambda_i^{(k)} f_i(x)\right)$$

//Obviously $x^{(k)}$ আগে calculate করতে হবে

rearranging the equations in order of calculation we arrive at the projected subgradient method:

Algorithm: Projected subgradient for dual problem

$$\boxed{x^{(k)} = \operatorname{argmin}_x \left(f_0(x) + \sum \lambda_i^{(k)} f_i(x)\right)}$$

Note that primal iterates do not necessarily satisfy $f_i(x_i) \leq 0 \big|_{i=1}^{m}$. whether

subgradient method: Algorithm: Projected subgradient for dual problem

$$x^{(k)} = \underset{x}{\text{argmin}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i^{(k)} f_i(x) \right)$$

$$\lambda^{(k+1)} = \left(\lambda^{(k)} - \alpha_k \left((-f_i(x^{(k)}))_{i=1}^{m} \right) \right)_+$$

elementwise

$$\forall_{i \in \{1,\dots,m\}} \quad \lambda_i^{(k+1)} = \left(\lambda_i^{(k)} + \alpha_k f_i(x^{(k)}) \right)_+$$

Note that primal iterates do not necessarily satisfy $f_i(x_i) \le 0 \big|_{i=1}^{m}$, whether

the dual iterates are always ≥ 0 as they are projection on $\mathbb{R}^n_+$.

Primal iterates become feasible in $k \to \infty$

if • resource i over utilized $(f_i(x^{(k)}) > 0)$, $\alpha_k > 0$ (by defn) $\therefore \alpha_k f_i(x^{(k)}) > 0$

$$\lambda_i^{(k+1)} = \underbrace{\left(\underbrace{\lambda_i^{(k)}}_{\geq 0} + \underbrace{\alpha_k f_i(x^{(k)})}_{>0} \right)}_{>0}{}_+ = \left(\lambda_i^{(k)} + \underbrace{\alpha_k f_i(x^{(k)})}_{>0} \right) > \lambda_i^{(k)}$$

, so price update increases the price for over consumption

• resource i usage is under utilized $(f_i(x^{(k)}) < 0)$ $\alpha_k \geq 0$, then $\alpha_k f_i(x^{(k)}) < 0$

$$\therefore \lambda_i^{(k+1)} = \left(\underbrace{\lambda_i^{(k)}}_{\geq 0} + \underbrace{\alpha_k f_i(x^{(k)})}_{<0} \right)_+ = \max\{ \lambda_i^{(k)} + \alpha_k f_i(x^{(k)}), 0 \} \leq \lambda_i^{(k)}$$

e.g. $\lambda_i^{(k)} = 3$, $\alpha_k f_i(x^{(k)}) = -5$

$\lambda_i^{(k+1)} = (3-5)_+ = (-2)_+ = 2 < \lambda_i^{(k)}$

$\lambda_i^{(k)} = 3$, $\alpha_k f_i(x^{(k)}) = -7$

$\lambda_i^{(k+1)} = (3-7)_+ = (-4)_+ = 0 < \lambda_i^{(k)}$

$\lambda_i^{(k)} = 0$, $\alpha_k f_i(x^{(k)}) = -3$
$\lambda_i^{(k+1)} = (0-3)_+ = (-3)_+ = 0$
} once $\lambda_i^{(k)}$ hits 0, it stays 0

Example:

$$\begin{pmatrix} \text{min} & \frac{1}{2} x^T \overset{\succ 0 \,\Rightarrow\, \text{strongly convex}}{P} x - q^T x \\ \text{s.t.} & \forall_{i \in \{1,\dots,n\}} \; x_i^2 \leq 1 \end{pmatrix}$$

$$\underset{\lambda \geq 0}{L(x,\lambda)} = \frac{1}{2} x^T P x - q^T x + \sum_{i=1}^{n} \lambda_i (x_i^2 - 1)$$

$$= \frac{1}{2} x^T P x - q^T x + \underbrace{\sum_{i=1}^{n} \lambda_i x_i^2} - \overbrace{\sum_{i=1}^{n} \lambda_i}^{1^T \lambda}$$

$x^T \text{diag}(\lambda_1,\dots,\lambda_n) x = \frac{1}{2} x^T \underbrace{\text{diag}(2\lambda_1,\dots,2\lambda_n)}_{\text{diag}(2\lambda)} x = \frac{1}{2} x^T \text{diag}(2\lambda) x$

$$= \frac{1}{2} x^T P x - q^T x + \frac{1}{2} x^T \text{diag}(2\lambda) x - 1^T \lambda$$

$$= \frac{1}{2} x^T (P + \text{diag}(2\lambda)) x - q^T x - 1^T \lambda \xrightarrow[\text{to find } x^*(\lambda)]{\nabla_x(\cdot) = 0} \underbrace{(P + \text{diag}(2\lambda))}_{\succ 0 \text{ so invertible}} x - q = 0 \to x = (P + \text{diag}(2\lambda))^{-1} q$$

$\because P \succ 0$, $\text{diag}(2\lambda) \succeq 0$
$\to P + \text{diag}(2\lambda) \succ 0$

$\therefore x^*(\lambda) = (P + \text{diag}(2\lambda))^{-1} q$

Projected subgradient algorithm will be: (Algorithm: Projected subgradient for dual problem (ctd))

$$x^{(k)} = (P + \text{diag}(2\lambda^{(k)}))^{-1} q$$

$$\lambda_i^{(k+1)} = \left(\lambda_i^{(k)} + \alpha_k \underbrace{f_i(x^{(k)})}_{(x_i^{(k)2} - 1)} \right)_+ = \left(\lambda_i^{(k)} + \alpha_k \left(x_i^{(k)2} - 1 \right) \right)_+$$

α can be determined by subgradient rules, or line search / backtracking as the dual function $g(\lambda) = \left(\frac{1}{2} x^T (P + \text{diag}(2\lambda)) x - q^T x - 1^T \lambda \right) \Big|_{x = x^*(\lambda)}$ is affine in λ hence differentiable.

primal iterates are not feasible, might violate $x_i^2 \leq 1 \leftrightarrow x_i \in [-1, 1]$

an nearly feasible construction is:

$$\tilde{x}_i^{(k)} = \begin{cases} 1, & x_i^{(k)} > 1 \\ -1, & x_i^{(k)} < 1 \\ x_i^{(k)}, & -1 \leq x_i^{(k)} \leq 1 \end{cases}$$

and then in $\lambda_i^{(k+1)}$ replace $x_i^{(k)}$ with $\tilde{x}_i^{(k)}$, so modified projected subgradient algorithm will be:

$$x^{(k)} = (P + \text{diag}(2\lambda^{(k)}))^{-1} q$$

$$\tilde{x}_i^{(k)} = \begin{cases} 1, & x_i^{(k)} > 1 \\ -1, & x_i^{(k)} < 1 \\ x_i^{(k)}, & -1 \leq x_i^{(k)} \leq 1 \end{cases}$$

$$\lambda_i^{(k+1)} = \left(\lambda_i^{(k)} + \alpha_k \underbrace{f_i(\tilde{x}^{(k)})}_{(\tilde{x}_i^{(k)2} - 1)} \right)_+ = \left(\lambda_i^{(k)} + \alpha_k \left(\tilde{x}_i^{(k)2} - 1 \right) \right)_+$$