Lagrangedualitet—repetition

$$[\mathrm{P}] \qquad v^* = \min \ f(x)$$
$$\text{då} \quad G_i(x) \le 0 \quad i = 1,\ldots,m \quad (1)$$
$$x \in X \qquad (2)$$

- $f, G_i : \Re^n \mapsto \Re$ kontinuerliga och $\emptyset \neq X \subseteq \Re^n$
- Relaxera bivillkoren (1) $\Rightarrow$ *subproblem* $\qquad (x(u) \in X(u) \subseteq X)$

$$g(u) = \min_{x \in X}\{f(x) + u^T G(x)\} = f[x(u)] + u^T G[x(u)]$$

- *Lagrangedualen*: maximera $g : \Re^m \mapsto \Re$ över $\Re^m_+$

$$[D] \qquad v_L = \max_{u \ge 0} g(u)$$

- *Svag dualitet*: $\quad g(u) \le v_L \le v^* \le f(x) \qquad$ då $u \ge 0$ och $x \in X$

Egenskaper hos den duala funktionen g

Sats: Antag att f och G_i, $i = 1,\ldots,m$, är kontinuerliga funktioner och att $X \subset \Re^n$ är icke-tom, sluten och begränsad. Då är g *ändlig* och *konkav*.

Bevis: *Ändlig:* (Weierstrass sats) För varje $u \ge 0$ är $f(x) + u^T G(x)$ kontinuerlig och antar därför sitt max och min på varje sluten och begränsad mängd (X).
Konkav: Låt $\hat{u}, \tilde{u} \in \Re^m_+$ och $\lambda \in [0,1]$.

$$\begin{aligned}
g(\lambda\hat{u} + [1-\lambda]\tilde{u}) &= \min_{x\in X}\left[f(x) + [\lambda\hat{u} + [1-\lambda]\tilde{u}]^T G(x)\right] \\
&= \min_{x\in X}\left\{\lambda\left[f(x) + \hat{u}^T G(x)\right] + (1-\lambda)\left[f(x) + \tilde{u}^T G(x)\right]\right\} \\
&\ge \lambda \min_{x\in X}\left\{f(x) + \hat{u}^T G(x)\right\} + (1-\lambda)\min_{x\in X}\left\{f(x) + \tilde{u}^T G(x)\right\} \\
&= \lambda g(\hat{u}) + (1-\lambda) g(\tilde{u})
\end{aligned}$$

Differentierbarhet

Sats: g är differentierbar i $u = \bar{u} \Leftrightarrow G(\bar{x})$ är konstant för alla $\bar{x} \in X(\bar{u}) \Rightarrow \nabla g(\bar{u}) = G(\bar{x})$

Följdsats: Om $X(\bar{u})$ består av en punkt (d.v.s. [$\mathrm{L_{sub}}$] har en unik lösning $x(\bar{u})$) så är g differentierbar i $u = \bar{u}$ och $\nabla g(\bar{u}) = G(x(\bar{u}))$.

En unik subproblemlösning fås t.ex. om $L(x,u)$ för fixt värde på u är *strikt konvex* m.a.p. x.

Konvexa problem med styrbarhet

Sats: Antag att problemet [P] är konvext. Om g är differentierbar i $u = u^*$ så är varje $x(u^*) \in X(u^*)$ en *global optimallösning* till [P].

Exempel 1 är ett sådant exempel där dualen är styrbar.

I Exempel 4 är g ej differentierbar i $u = u^*$; därför är inte varje punkt i $X(u^*)$ optimal (eller ens tillåten) i [P]; saknar styrbarhet.

LP-problem och linjära heltalsproblem saknar vanligen styrbarhet.

Hur gör man för att maximera g över $\Re^m_+$?

- Om g är differentierbar kan man använda gradientmetoder, t.ex. brantaste lutning (anpassad för bivillkoren $u \ge 0$), för att lösa [D].
- Vad gör vi om g ej är differentierbar, vilket är fallet t.ex. för de flesta kombinatoriska optimeringsproblem?
- Vi behöver en generalisering av gradientbegreppet.
- För en *differentierbar, konkav* funktion $h : \Re^m \mapsto \Re$ gäller

$$h(x) \le \underbrace{h(\bar{x}) + \nabla h(\bar{x})^T(x - \bar{x})}_{\text{stödjande hyperplan till } h \text{ i } \bar{x}} \qquad \text{för alla} \quad x \in \Re^m$$

Subgradienter

Definition: En vektor ξ är en *subgradient* till den *konkava* funktionen $h : \Re^m \mapsto \Re$ i $\bar{x} \in \Re^m$ om

$$h(x) \le h(\bar{x}) + \xi^T(x - \bar{x}) \quad \text{för alla} \quad x \in \Re^m$$

Definition: Mängden av alla subgradienter till h i $\bar{x} \in \Re^m$ bildar *subdifferentialen*

$$\partial h(\bar{x}) = \left\{\xi \in \Re^m \,\middle|\, h(x) \le h(\bar{x}) + \xi^T(x - \bar{x}) \quad \text{för alla} \quad x \in \Re^m\right\}$$

En subgradient är *lutningen* hos ett stödjande hyperplan till h i $\bar{x}$

En subgradient är *normal till nivåkurvan* för h i $\bar{x}$

BILDER!

Exempel A

Låt $h(x) = \min\{h_1(x), h_2(x)\}$ där $h_1(x) = 4 - |x|$ och $h_2(x) = 4 - (x-2)^2$

$$h(x) = \begin{cases} 4 - x, & 1 \le x \le 4 \\ 4 - (x-2)^2, & x \le 1, \ x \ge 4 \end{cases}$$

Funktionen h är *icke-differentierbar* i punkterna $x = 1$ och $x = 4$, därför att där har den *inget unikt stödplan*

$$\partial h(x) = \begin{cases} \{-1\}, & 1 < x < 4 \\ \{4 - 2x\}, & x < 1, \ x > 4 \\ [-1, 2], & x = 1 \\ [-4, -1], & x = 4 \end{cases}$$

Subdifferentialen är ett intervall i var och en av de icke-differentierbara punkterna

Subdifferentialen

Sats: Om $\partial h(\bar{x})$ består av ett element så är h differentierbar i $\bar{x}$ och $\partial h(\bar{x}) = \{\nabla h(\bar{x})\}$.

Sats: $\partial h(\bar{x})$ är en sluten, begränsad och konvex mängd.

Hur kan subdifferentialen skrivas om h är en min-funktion av typen i Exempel A?

Sats: Låt $h(x) = \min_{i=1,\ldots,m} h_i(x)$, där h_i är konkav och differentierbar, $i = 1,\ldots,m$. Då är h en konkav funktion.
Låt $\mathcal{I}(\bar{x}) \subseteq \{1,\ldots,m\}$ definieras av $h(\bar{x}) = h_i(\bar{x})$ för $i \in \mathcal{I}(\bar{x})$ och $h(\bar{x}) < h_i(\bar{x})$ för $i \notin \mathcal{I}(\bar{x})$ (de aktiva segmenten i $\bar{x}$).
Då är $\partial h(\bar{x})$ det *konvexa höljet* av $\{\nabla h_i(\bar{x}) \mid i \in \mathcal{I}(\bar{x})\}$, d.v.s.

$$\partial h(\bar{x}) = \left\{\xi = \sum_{i \in \mathcal{I}(\bar{x})} \lambda_i \nabla h_i(\bar{x}) \,\middle|\, \sum_{i \in \mathcal{I}(\bar{x})} \lambda_i = 1; \ \lambda_i \ge 0, i \in \mathcal{I}(\bar{x})\right\}.$$

Optimalitetsvillkor

För en differentierbar, konkav funktion h gäller:

$$x^* \text{ löser } \max_{x \in \Re^n} h(x) \qquad \Leftrightarrow \qquad \nabla h(x^*) = 0$$

Sats: Låt h vara konkav och icke-differentierbar. Då gäller

$$x^* \text{ löser } \max_{x \in \Re^n} h(x) \qquad \Leftrightarrow \qquad 0 \in \partial h(x^*)$$

Bevis: Antag att $0 \in \partial h(x^*) \Rightarrow h(x) \le h(x^*) + 0^T(x - x^*)$ för alla $x \in \Re^n$, d.v.s., $h(x) \le h(x^*)$ för alla $x \in \Re^n$.
Antag att x^* löser $\max_{x \in \Re^n} h(x) \Rightarrow$
$h(x) \le h(x^*) = h(x^*) + 0^T(x - x^*)$ för alla $x \in \Re^n$, d.v.s., $0 \in \partial h(x^*)$

Exempel A: $0 \in \partial h(1) \Rightarrow x^* = 1$

För optimeringsproblem med bivillkor generaliseras KKT-villkoren:

$$x^* \text{ löser } \max_{x \in X} h(x) \qquad \Leftrightarrow \qquad \partial h(x^*) \cap N_X(x^*) \neq \emptyset$$

där $N_X(x^*)$ är koniska höljet av bindande bivillkorens normaler i x^*.

Subgradientalgoritmen

- I differentierbara fallet gäller att $\nabla h(\bar{x})$ pekar i den riktning som (lokalt) ökar värdet på h snabbast
- Brantaste lutnings-algoritmen bygger på detta faktum
- Om h inte är differentierbar och vi istället vill utnyttja subgradienter som stegriktningar, så kan det hända att värdet på h *minskar* i en subgradientriktning.
- En annan egenskap hos differentierbara, konkava funktioner är att gradienten $\nabla h(\bar{x})$ alltid *pekar in i ett halvrum som innehåller optimum*, eller, $x^* \in \{x \in \Re^n \mid \nabla h(\bar{x})^T(x - \bar{x}) \geq 0\}$

Subgradienten pekar in i ett halvrum som innehåller optimum

Sats: Låt $\xi \in \partial h(\bar{x})$ och låt X^* vara mängden av optimala lösningar till $\max_{x \in \Re^n} h(x)$. Då gäller $X^* \subseteq \{x \in \Re^n \mid \xi^T(x - \bar{x}) \geq 0\}$

Bevis: Per definition: $h(x) \leq h(\bar{x}) + \xi^T(x - \bar{x})$ för alla $x \in \Re^n$.
Låt $x^* \in X^* \Rightarrow h(x) \leq h(x^*)$ för alla $x \in \Re^n$. Speciellt gäller att $h(\bar{x}) \leq h(x^*) \Rightarrow h(\bar{x}) \leq h(x^*) \leq h(\bar{x}) + \xi^T(x^* - \bar{x}) \Rightarrow \xi^T(x^* - \bar{x}) \geq 0$

- Ett litet steg i riktning ξ för oss alltså närmre optimum.
- Dock är det inte säkert att värdet på h ökar i varje sådant steg.
- Idé: Använd subgradienten ξ på samma sätt som ∇h i brantaste lutnings-algoritmen
- Problem: Inte alla subgradienter är ascentriktningar $\Rightarrow$ linjesökning för bestämning av steglängden fungerar inte.

Det finns alltid subgradienter till h i $\bar{x}$ som är ascent-riktningar, men de är dock inte alltid lätta att finna (kräver mycket mera arbete än att finna bara en subgradient)

Subgradientalgoritmen

$$x^0 \in \Re^n, \quad x^{k+1} = x^k + t_k \cdot \xi^k$$

där $\xi^k \in \partial h(x^k)$ är en godtycklig subgradient och $t_k > 0$ steglängden.

- Hur hanteras bivillkor?
- Hur väljs steglängden t_k ?
- Hur kan man avbryta subgradientalgoritmen?
- När är algoritmen konvergent, d.v.s., när gäller $\{x^k\} \to x^*$?
- Hur kan man använda algoritmen för att lösa [D]?

Bivillkor

Om vi vill lösa $\min_{x \in X} h(x)$ där $\emptyset \neq X \subset \Re^n$ är konvex och h är konkav, så blir algoritmen:

$$x^0 \in X, \quad x^{k+1} = \mathtt{Proj}_X\left[x^k + t_k \cdot \xi^k\right]$$

där $\mathtt{Proj}_X[\hat{x}]$ är lösningen till projektionsproblemet:

$$\min_{x \in X} \|x - \hat{x}\|_2$$

(Euklidisk projektion på den tillåtna mängden.)

Steglängder

Sats: Låt X^* vara mängden av optimallösningar och h^* det optimala målfunktionsvärdet. Låt $x^k \in X \setminus X^*$ och $\xi^k \in \partial h(x^k)$. Om steglängden t_k väljs i det öppna intervallet $\left(0, 2\frac{h^* - h(x^k)}{\|\xi^k\|_2^2}\right)$ så gäller att $\|x^{k+1} - x^*\|_2 < \|x^k - x^*\|_2$ för varje $x^* \in X^*$.

$\Rightarrow$ **Steglängdsformel:** $t_k = \gamma_k \frac{h^* - h(x^k)}{\|\xi^k\|_2^2}$, där $\gamma_k \in (0, 2)$.

Oftast är h^* inte känd, men kan ersättas av $\bar{h}^k \geq h^*$, som uppdateras.

Konvergens

Sats: Låt $x^0 \in X$ och sekvensen $\{x^k\}$ ges av subgradientalgoritmen med steglängder enligt $t_k = \gamma_k \frac{h^* - h(x^k)}{\|\xi^k\|_2^2}$, där $0 < \varepsilon_1 \leq \gamma_k \leq \varepsilon_2 < 2$. Om sekvensen $\{\xi^k\}$ är begränsad så gäller att $\{x^k\} \to x^* \in X^*$.

Andra steglängdsval är möjliga, t.ex., divergent serie:
$t_k > 0, \forall t; \quad \lim_{k \to \infty} t_k = 0; \quad \sum_{k=0}^{\infty} t_k = \infty \quad$ (ev även $\sum_{k=0}^{\infty} t_k^2 < \infty$)

Tillämpning på Lagrangedualen

$$[D] \qquad v_L = \max_{u \geq 0} g(u)$$

$$g(u) = \min_{x \in X}\{f(x) + u^T G(x)\} = f[x(u)] + u^T G[x(u)]$$

En subgradient till g i $\bar{u}$ ges av $G[x(\bar{u})]$, d.v.s., lösningen till subproblemet insatt i de relaxerade bivillkoren.

Subgradientalgoritmen:

$$u^0 \geq 0, \quad u^{k+\frac{1}{2}} = u^k + t_k \cdot G[x(u^k)], \quad u_i^{k+1} = \max\left\{u_i^{k+\frac{1}{2}}; 0\right\}$$

Ofta kompletteras algoritmen med en primal tillåtenhets-heuristik: Finn en tillåten lösning till [P] genom att manipulera $x(u^k)$ lite $\to$ $\hat{x}^k \in X \Rightarrow$ en kandidat till optimum och en övre gräns: $f(\hat{x}^k) \geq v^*$.

Möjliga förbättringar

- Välj subgradienten ξ^k mer omsorgsfullt, så att målfunktionsvärdet förbättras. Dock; oftast mycket mer beräkningskrävande. (Bundle methods)
- Vrid ξ^k ungefär som i Newton-metoder. (Space dilation methods)
- Gör ξ^k (lokalt) tillåten genom att projicera på den tillåtna mängden. (Subgradient-projektion)

Linjära heltalsproblem: Styrkan hos en Lagrangedual

Jämför med en kontinuerlig relaxering (LP-relaxering):

$$\begin{array}{rlcrl} v_{LP} = \min & c^T x & \leq & v^* = \min & c^T x \\ \text{då} & Ax \leq b & & \text{då} & Ax \leq b \\ & Dx \leq d & & & Dx \leq d \\ & x \in \Re_+^n & & & x \in Z_+^n \end{array}$$

Låt $X = \{x^1, x^2, \ldots, x^K\}$ beteckna mängden $X = \{x \in Z_+^n \mid Ax \leq b\}$.

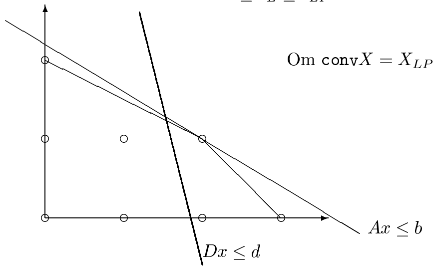
$$\begin{aligned} \Rightarrow \; v_L &= \max_{\lambda \geq 0}\left(\min_{x \in X}\left[c^T x + \lambda^T(Dx - d)\right]\right) \\ &= \max_{\lambda \geq 0}\left(\min_{k=1,\ldots,K}\left[c^T x^k + \lambda^T(Dx^k - d)\right]\right) \end{aligned}$$

$$
\begin{aligned}
&= \max_{\lambda\geq 0,\theta\in\Re} \left\{\theta \,\middle|\, \theta-(Dx^k-d)^T\lambda \leq c^Tx^k, \quad k=1,\dots,K\right\} \\
&\quad \text{/ LP-dualvariabler, } y_k \Rightarrow \text{/} \\
&= \min \quad \sum_{k=1}^K (c^Tx^k)y_k \\
&\quad \text{då} \quad \sum_{k=1}^K y_k = 1 \\
&\quad\quad \sum_{k=1}^K (Dx^k-d)y_k \leq 0 \Leftrightarrow D\underbrace{\sum_{k=1}^K x^k y_k}_{\in \mathtt{conv}X} \leq d\underbrace{\sum_{k=1}^K y_k}_{=1} \\
&\quad\quad y_k \geq 0, \quad k=1,\dots,K \\
&= \min \quad c^Tx \\
&\quad \text{då} \quad Dx\leq d, \quad x\in \mathtt{conv}X
\end{aligned}
$$

Styrkan hos en Lagrangedual

Eftersom $X \subseteq \mathtt{conv}X \subseteq X_{LP} = \{x\in\Re^n_+ \mid Ax\leq b\}$ så gäller

$$v^* \geq v_L \geq v_{LP}$$



Om $\mathtt{conv}X = X_{LP} \Rightarrow v_L = v_{LP}$

Heltalsegenskap

- Om $\min_{x\in X_{LP}} g^Tx = \min_{x\in\mathtt{conv}X} g^Tx$, för alla $g\in\Re^n$ d.v.s., om subproblemet har *heltalsegenskap*, så gäller: $v_L = v_{LP}$.
- Annars är v_L en *bättre* optimistisk uppskattning av v^* än vad v_{LP} är. D.v.s. $v_L \geq v_{LP}$.
- Heltalsegenskap $\underbrace{\Leftrightarrow}_{\text{oftast}}$ lätt att lösa, låg komplexitet
- Låg komplexitet hos subproblemet "⇒" Dålig uppskattning
- Högre komplexitet hos subproblemet "⇒" Bättre uppskattning

⇒ Subproblemet skall *inte* väljas så att det blir *för lätt* att lösa