Corso:

Metodi matematici e statistici e Ricerca Operativa

LM Ingegneria dei Trasporti e della Logistica, Università di Genova, 2009/10
Co-docenza, con Marcello SANGUINETI

Fioravante Patrone [http://www.diptem.unige.it/patrone/default.htm]
Sezione Metodi e Modelli Matematici, DIPTEM
Facoltà di Ingegneria
Università di Genova

E' una versione provvisoria (e penso resterà tale per sempre).
Consultare la data/ora dell'ultimo aggiornamento (vedi in fondo).

Solo i colti amano imparare; gli ignoranti preferiscono insegnare - Edouard Le Berquier

Da un mail ricevuto:
significa che, salvo imprevisti inattesi (per definizione, forse), io ci sono eccome.

Gi 5 novembre 2009, 3h-3h:
Introduzione alla "mia" parte di corso: programma previsto.
Massimi e minimi, definizioni.
Parliamo di max (per min basta cambiare segno)
max per $f: A \to B$ con A e B insiemi qualsiasi non ha senso. Il "problema" è su B, perché per parlare di max ho bisogno che su B ci sia un ordine. Per farla breve, prendiamo $B = \mathbb{R}$.
Abbiamo quindi $f: A \to \mathbb{R}$. Con A insieme qualsiasi.
Def di punto di max globale e di (valore) massimo globale (dedotta da quella di punto di max).
Si può anche seguire strada opposta, partendo da max di $f(A)$.
Punto di max stretto.
Non ha senso parlare di max locale. Dobbiamo poter parlare di intorni. Prendiamo $A \subseteq \mathbb{R}^n$
$f: A \subseteq \mathbb{R}^n \to \mathbb{R}$.
Def di max locale: c'è intorno I (sferico, se vogliamo) di $\overline{x}$ t.c. $\overline{x}$ sia punto di max (globale) per $f|(I \cap A)$
Caso particolare $A \subseteq \mathbb{R}$. C'è $\delta > 0$ tale che, posto $I =]\overline{x} - \delta, \overline{x} + \delta[$, il punto $\overline{x}$ è punto di max globale per f ristretta a $I \cap A$.

Nota: sarebbe carino, soprattutto dal punto di vista applicativo, potere essere in grado di stimare quali δ "vanno bene", e non limitarsi semplicemente a dire: ci sarà un $\delta > 0$ che "va bene".
Vediamo un esempio semplice in cui si può ottenere una stima di δ. Proprio per rendere molto semplice l'esempio, assumerò che f sia definita su tutto $\mathbb{R}$, che abbia derivata terza su tutto $\mathbb{R}$ e che esista $M > 0$ t.c. $|f'''(x)| \le M$ per ogni $x \in \mathbb{R}$ (è soprattutto questa limitazione, assunta valida su tutto $\mathbb{R}$ che rende l'esempio non troppo applicabile. Ma in realtà il discorso si può fare anche su un intervallo, io mi metto su tutto $\mathbb{R}$ per dare una idea di come si possa stimare il summenzionato δ).
Allora, supponiamo che sia $f'(\overline{x}) = 0$ e $f''(\overline{x}) = \beta < 0$.
Scriviamo la formula di Taylor con il resto di Lagrange: esiste $\xi \in \mathbb{R}$ t.c.:
$f(x) = f(\overline{x}) + f'(\overline{x}) \cdot (x - \overline{x}) + \frac{1}{2} f''(x - \overline{x})^2 + \frac{1}{6} f'''(\xi) \cdot (x - \overline{x})^3$.
Allora:
$f(x) - f(\overline{x}) = 0 + \frac{\beta}{2} \cdot (x - \overline{x})^2 + \frac{1}{6} f'''(\xi) \cdot (x - \overline{x})^3 \le$
$\frac{\beta}{2} \cdot (x - \overline{x})^2 + \frac{1}{6} M |x - \overline{x}|^3$.
Per avere $f(x) - f(\overline{x}) \le 0$ è allora sufficiente avere $\frac{\beta}{2} \cdot (x - \overline{x})^2 + \frac{1}{6} M |x - \overline{x}|^3 \le 0$.
Abbiamo $\frac{\beta}{2} \cdot (x - \overline{x})^2 + \frac{1}{6} M |x - \overline{x}|^3 \le 0$ se e solo se $\beta + \frac{1}{3} M |x - \overline{x}| \le 0$, il che è equivalente a $|x - \overline{x}| \le -\frac{3\beta}{M}$. Insomma, possiamo prendere $\delta = -\frac{3\beta}{M}$ (ricordiamo che $\beta < 0$).

CN di max "libero" per funzioni di una variabile.
Proviamo a ricostruire il teorema. Supponiamo di avere un punto di max locale (le CN che troveremo per un punto di max locale valgono, *a fortiori*, per un max globale).
Che f sia derivabile nel punto di max loc non lo possiamo ottenere come tesi, vedi esempio di $|x|$.
Proviamo a dimostrare il teorema, per funzioni di una variabile. E' data $f: A \to \mathbb{R}$, con $A \subseteq \mathbb{R}$. E $\overline{x}$ è p.to di max loc.
Per la dim, tre idee:
1. considerare separatamente $\overline{x} < x$ e $\overline{x} > x$.
2. scrivere il rapporto incrementale e notare che ha segno definito (è ≤ 0 se $\overline{x} < x$) .
3. usare teorema di permanenza del segno
Si ha allora $f'(\overline{x}) \le 0$, "arrivando da destra". Se si arriva "da sinistra", $f'(\overline{x}) \ge 0$. Allora $f'(\overline{x}) = 0$. Ma attenzione, stiamo dimenticando cose importantissime!!!

Vediamo i dettagli.
Per arrivare a $f'(\overline{x}) = 0$, aggiungiamo la condizione che $\overline{x}$ sia interno ad A (la condizione davvero necessaria è che $\overline{x}$ sia punto di accumulazione sia da destra che da sinistra. Se lo è solo da una parte, possiamo comunque scrivere la condizione "unilaterale").
Sono comunque interessanti anche le condizioni "unilaterali" che, nel caso di un intervallo, si hanno negli estremi (se $A = [a, b]$, ed a è massimo, $f'(a) \le 0$; se b è massimo, $f'(b) \ge 0$).

Una curiosità: perché è utile la condizione $f'(x) = 0$? Perché ci permette di cercare punti di max locale risolvendo una equazione nella variabile x. Mentre se vogliamo utilizzare direttamente la definizione, dobbiamo trovare $\delta > 0$ t.c.:
$f(\overline{x}) \ge f(x)$ per ogni x t.c. $|x - \overline{x}| < \delta$.
Ovvero, risolvere una famiglia di disequazioni parametrizzate da $\overline{x}$, e poi sulla base di quanto trovato cercare se si trova $\delta > 0$ in modo da soddisfare la condizione sopra riportata. In effetti, sembra essere una procedura un po' più laboriosa. Anche se ha il vantaggio di consegnarci i max locali, mentre la ben nota $f'(x) = 0$ è solo una condizione necessaria.

NOTA: la strada descritta qui non è stata seguita a lezione. La lascio per chi sia interessato.
Generalizzazione ad $\mathbb{R}^n$: sia $\overline{x}$ un p.to di max locale per $f: A \to \mathbb{R}$, con $A \subseteq \mathbb{R}^n$. Allora, dato un versore $v \in \mathbb{R}^n$ (si potrebbe anche prendere un vettore, basta che $v \neq 0$), $\lim_{t \to 0^+} \frac{f(x+tv)-f(x)}{t} \leq 0$ se ha senso. Ovviamente, $f(x+tv)$ deve essere definita per "$t > 0$ sufficientemente piccoli". Vale a dire, abbiamo bisogno che il punto $\overline{x}$ sia di accumulazione per $A \cap S_{\overline{x},v}$, dove $S_{\overline{x},v}$ indica la semiretta uscente da $\overline{x}$, individuata dal versore v.
Quindi: $\frac{\partial f}{\partial v^+}(\overline{x}) \leq 0$ se ha senso. [Con $\frac{\partial f}{\partial v^+}(x)$ indichiamo il limite scritto sopra: $\lim_{t \to 0^+} \frac{f(x+tv)-f(x)}{t}$].

In particolare, se f è parzialmente derivabile ed $\overline{x}$ è interno ad A, per ogni $i \in \{1, \dots, n\}$ si ha $\frac{\partial f}{\partial x_i}(\overline{x}) = 0$.

Questo fatto deriva facilmente dalla osservazione che (in due variabili, con n variabili i conti sono simili):
--- $\frac{\partial f}{\partial x^+}(\overline{x}) = \frac{\partial f}{\partial v^+}(\overline{x}) \leq 0$, prendendo come v il versore $(1, 0)$.
--- mentre $\frac{\partial f}{\partial x^-}(\overline{x}) = -\frac{\partial f}{\partial w^+}(\overline{x})$, dove w è il versore $(-1, 0)$. Essendo $\frac{\partial f}{\partial w^+}(\overline{x}) \leq 0$, otteniamo che $\frac{\partial f}{\partial x^-}(\overline{x}) \geq 0$.
NOTA FINE della parte non fatta a lezione.

Per ottenere le CN necessarie per una funzione $f: A \to \mathbb{R}$, con $A \subseteq \mathbb{R}^n$, possiamo fare le considerazioni seguenti.
Prima di tutto, prendiamo x_0 p.to di max locale per f. Cioè esiste $\delta_1 > 0$ t.c. $f(x_0) \geq f(x)$ per ogni $x \in I_{\delta_1}(x_0) \cap A$.
$I_\delta(x_0)$ indica l'intorno di x_0 di raggio δ, ovvero l'insieme di tutti i punti di $\mathbb{R}$ che hanno distanza euclidea da x_0 minore strettamente di δ.
Supponiamo che x_0 sia *interno* ad A, ovvero che esista $\delta_2 > 0$ t.c. $I_{\delta_2}(x_0) \subseteq A$.
Scegliamo $\delta = \min\{\delta_1, \delta_2\}$. Allora, x_0 è punto di massimo locale (e globale) per f ristretta a $I_\delta(x_0)$. Non solo, ma il "diametro" passante per x_0 e parallelo all'asse delle x_1 è tutto contenuto in A. Possiamo allora applicare il teorema provato per le funzioni di una variabile alla restrizione di f a questo "diametro". Ne segue che la derivata prima di questa funzione sarà nulla. Ma essa coincide con la derivata parziale di f rispetto alla variabile x_1. Visto che un discorso analogo si può fare per tutte le derivate parziali, otteniamo che $\frac{\partial f}{\partial x_i}(x_0) = 0$ per ogni $i \in \{1, \dots, n\}$.
In dettaglio, la funzione di una variabile che abbiamo considerato, è definita così: $\varphi(t) = f\left(t, x_{0_2}, \dots, x_{0_n}\right)$. La funzione φ è definita (almeno) per $t \in (x_0 - \delta, x_0 + \delta)$.

Per lavorare "tranquillamente" useremo delle ipotesi più forti su f, tipo la differenziabilità.
Ma la verifica diretta della differenziabilità può essere ostica. Se applicabile, possiamo sfruttare il fatto che, se $f \in \mathcal{C}^1(A)$, allora è differenziabile in ogni punto di A
Attenzione, però, che questo risultato "richiede" che l'insieme A sia aperto (ho messo le virgolette a *richiede*perché non è obbligatoria. E' una questione di opportunità: la definizione di $\mathcal{C}^1(A)$, se A non è aperto, è molto laboriosa).

DOCUMENTI:

.
ESERCIZI:
.

Ma 10 novembre 2009, 2h-5h:
Un commento: quando le condizioni del primo ordine sono sufficienti? Per le funzioni concave!
Data $f \in \mathcal{C}^1(A)$, con A aperto e convesso di $\mathbb{R}^n$, se f è concava allora, per $\overline{x} \in A$:
$\nabla f(\overline{x}) = 0$ se e solo se $\overline{x}$ è punto di max assoluto per f.
Lo si dimostra usando la caratterizzazione differenziale della concavità: $f \in \mathcal{C}^1(A)$, con A aperto, è concava se e solo se per ogni $\overline{x}, x \in A$ si ha $f(x) \leq f(\overline{x}) + \nabla f(\overline{x}) \cdot (x - \overline{x})$.

Ci sono poi le condizioni del secondo ordine.
L'idea generale è questa. Ho f. La approssimo con una funzione "semplice" (ad esempio un polinomio). Guardo sotto quali condizioni questa funzione approssimante ha max. Cerco di provare se queste condizioni possano essere necessarie perché la f abbia max (o sufficienti, magari "rafforzandole un poco": ad esempio, chiedendo che la funzione approssimante abbia un punto di max *stretto*)
Supponiamo che sia $f \in \mathcal{C}^2(A)$ (è importante sottolineare come l'idea di approssimazione sia una idea di carattere generale, usata in molti contesti).
Noi approssimeremo f con il polinomio di Taylor (del secondo ordine): $f(\overline{x}) + \sum_{i=1,\dots,n} \frac{\partial f}{\partial x_i}(\overline{x})(x_i - \overline{x}_i) + \sum_{i,j=1,\dots,n} \frac{\partial^2 f}{\partial x_i x_j}(\overline{x})(x_i - \overline{x}_i)(x_j - \overline{x}_j)$
.
Per provare quanto affermeremo, occorre naturalmente provare che il "resto" della formula di Taylor può essere davvero trascurato, ai nostri fini. Questo si ottiene esprimendo tale esito nella forma di Peano e sfruttando il fatto che va a zero più rapidamente del quadrato della distanza di x da $\overline{x}$.

La parte su cui concentrare l'attenzione è la forma quadratica (detta forma quadratica hessiana o, semplicemente, forma hessiana) :
$\sum_{i,j=1,\dots,n} \frac{\partial^2 f}{\partial x_i x_j}(\overline{x})(x_i - \overline{x}_i)(x_j - \overline{x}_j)$. Ci aspettiamo **grosso modo** che essa abbia max in $x = \overline{x}$, se $\overline{x}$ è un punto di massimo per f.

Per comodità chiamiamo $\Phi(h) = \sum_{i,j=1,\dots,n} \frac{\partial^2 f}{\partial x_i x_j}(\overline{x}) h_i h_j$ e vediamo se Φ ha max in 0 .
Chiaramente Φ ha max in 0 se e solo se $\Phi(h) \leq 0$ per ogni $h \in \mathbb{R}^n$ (si noti che $\Phi(0) = 0$).
se $\Phi(h) \leq 0$ per ogni $h \in \mathbb{R}^n$, si dice che Φ è semidefinita negativa..

Come verificare se Φ è semidefinita negativa? Con la matrice hessiana: $(H_{ij})_{i,j=1}^{n} = \frac{\partial^2 f}{\partial x_i x_j}(\bar{x})$. E' una matrice simmetrica ($f \in \mathcal{C}^2(A)$ e quindi vale il teorema di Schwartz). Quindi diagonalizzabile (e gli autovalori sono tutti reali). Si dimostra che Φ sarà semidefinita negativa se e solo se tutti i suoi autovalori saranno minori o uguali a zero.
Invece, la condizione $\Phi(h) < 0$ per ogni $h \neq 0$, cioè se Φ abbia in 0 un punto di max stretto, è una condizione sufficiente perché $\bar{x}$ sia un punto di max locale (ammesso che sia anche $\nabla f(\bar{x}) = 0$).
Se $\Phi(h) > 0$ per ogni $h \neq 0$, diciamo che Φ è definita negativa.
E Φ è definita negativa se e solo se ogni suo autovalore è strettamente negativo
Notare che se Φ ha un autovalore strettamente positivo ed uno strettamente negativo, il punto $\bar{x}$ non potrà essere né punto di minimo né punto di massimo. Anzi, possiamo affermare con certezza che si tratta di un punto di sella.

NOTA:
Le condizioni sufficienti del secondo ordine, se sono soddisfatte nel punto $\bar{x}$, ci garantiscono che la nostra funzione ha un punto di max locale in $\bar{x}$. Cioè, il punto $\bar{x}$ è un punto di massimo globale per la funzione f ristretta ad un opportuno intorno di $\bar{x}$. Ma, come già detto in precedenza, non abbiamo nessuna informazione su come sia grande questo intorno. Una stima sull'ampiezza massima dell'intorno che si può prendere può essere molto utile per le applicazioni. E' possibile ottenerla, se si hanno informazioni sulle derivate terze ed, in particolare, se si hanno a disposizione opportune maggiorazioni per tali derivate, come abbiamo visto.

NOTA:
La idea di approssimare f con una funzione ad essa "vicina" e allo stesso tempo "semplice" va ben al di là del caso della ricerca di CN e CS per punti di max locali. Non solo, ci sono diverse idee per approssimare una funzione, che sono utili in contesti diversi. Mi limito a citare un caso: la interpolazione polinomiale. Essa ha un ruolo diverso dalla approssimazione ottenuta dalla formula di Taylor.

DOCUMENTI:
Per chi fosse interessato: cosa può capitare se si cercano massimi (o minimi) locali di una funzione composta. NEW 31 mag 2010.
ESERCIZI:

.
ESEMPI:
Un paio di esempi semplici ma significativi sono:
$f(x, y) = x^2 + y^3$ ha come unico punto critico (cioè, punto in cui si annulla il gradiente) solo $(0, 0)$. La forma quadratica è $\Phi(h, k) = h^2$, ed è ovviamente semidefinita positiva ma non è definita positiva (si annulla in $(h, k) = (0, 1)$. Quindi è soddisfatta la CN ma non la CS. E il punto critico non è né punto di max locale né di min locale. Si noti che non è neanche un punto si sella.
$f(x, y) = x^2 + y^4$ ha come unico punto critico (cioè, punto in cui si annulla il gradiente) solo $(0, 0)$. La forma quadratica è $\Phi(h, k) = h^2$, identica a quella della precedente funzione, ovviamente semidefinita positiva ma non è definita positiva (si annulla in $(h, k) = (0, 1)$. Quindi è soddisfatta la CN ma non la CS. In questo caso il punto critico è punto di min locale.

Ve 13 novembre 2009, 3h-8h:
Problemi di max vincolato. Non è netta la distinzione fra pb libero e vincolato. Un pb libero lo possiamo vedere come vincolato e viceversa. Dipende da come ci sono forniti i dati, ed anche da quali siano le tecniche più opportune da applicare, nello specifico contesto in cui ci si trova.
Vincoli di uguaglianza: $h_i(x) = 0$, con $h_i : A \to \mathbb{R}$, $i \in U$. L'insieme U è un insieme finito di indici, può anche essere vuoto, il che sta ad indicare che non abbiamo vincoli di uguaglianza...
Vincoli di disuguaglianza: $g_j(x) \leq 0$, con $g_j : A \to \mathbb{R}$, $j \in D$. Anche D è un insieme finito di indici, e se è vuoto non abbiamo vincoli di disuguaglianza.
Un po' strano che tutte le funzioni siano definite sullo stesso A, ma non è condizione restrittiva.
Faremo poi sempre l'ipotesi che A sia un aperto di $\mathbb{R}^n$.

Il teorema di Weierstrass garantisce esistenza di p.to di massimo (globale). Ipotesi: $f : E \to \mathbb{R}$, $E \subseteq \mathbb{R}^n$ chiuso e limitato, f continua. In queste ipotesi siamo certi che f abbia massimo (e quindi almeno un punto di max) e, anche, se ci servisse, che abbia minimo.
C'è anche teorema di Weierstrass generalizzato: se E non è limitato (pur essendo chiuso), basta che $\lim_{||x|| \to \infty} f(x) = -\infty$, per essere certi della esistenza di un punto di max globale. (Se ci interessassero i min, dovremmo sostituire la condizione: $\lim_{||x|| \to \infty} f(x) = -\infty$ con $\lim_{||x|| \to \infty} f(x) = +\infty$)

Una applicazione consueta prevede $E = \left\{x \in \mathbb{R}^n : h_i = 0 \;\; per \;\; i \in U, g_j \geq 0 \;\; per \;\; j \in D\right\}$. Se le funzioni h_i e g_j sono definite su tutto $\mathbb{R}$ e sono continue, l'insieme E è chiuso.
Ricordo che un insieme è chiuso se e solo se il suo complementare è un insieme aperto. Ed è facile dimostrare che l'insieme delle $x \in \mathbb{R}^n$ per cui, ad esempio, è $h_i(x) \neq 0$, risulta essere un insieme aperto (basta usare il teorema di permanenza del segno). Allora l'insieme per cui $h_i(x) = 0$ è chiuso in quanto complementare di un aperto. Similmente si ragione per le g_j.
Ultimo tassello necessario: osservare che la intersezione di un numero finito di insiemi chiusi è un insieme chiuso (si prova infatti facilmente che la unione di un numero finito di insiemi aperti è un aperto, semplicemente usando la definizione. E poi si passa alla intersezione dei complementari, usando De Moivre).

Massimi e minimi vincolati (al solito, parleremo solo di max).
Abbiamo già visto il "setting". Abbiamo $A \subseteq \mathbb{R}^n$, e sono date f, h_i, con $i = 0, ..., m$ e g_j, $j = 0, ..., r$. Tutte definite su A e a valori in $\mathbb{R}$
Cerchiamo p.ti di max (locale, globale) per $f|\ E$ dove E è l'insieme dei punti che soddisfano il vincolo (o, brevemente, il vincolo). Cioè è l'insieme dei punti $x \in A$ t.c. $h_i(x) = 0$ e $g_j(x) \leq 0$ per ogni i e j.
Supponiamo D non vuoto e D vuoto. Cioè, *abbiamo* vincoli di uguaglianza e *solo* quelli.
Supponiamo f e le h_i di classe $\mathcal{C}^1(A)$. Ricordare che serve A aperto, per parlare di $\mathcal{C}^1(A)$
Se $x^\star$ è punto di massimo vincolato, abbiamo la condizione detta "dei moltiplicatori di Lagrange".
Occorre assumere anche che i vettori $\nabla h_i(x^\star)$ siano *linearmente indipendenti*.
Se questa condizione non è soddisfatta, non abbiamo garanzia che la tesi del teorema dei moltiplicatori di Lagrange valga. Ciò significa, in parole povere, che punti (in E, ovvero che stiano sul vincolo!) nei quali i vettori $\nabla h_i(x^\star)$ non sono linearmente indipendenti andranno sempre

messi nella lista dei sospettabili di essere punti di max locale.

Allora esiste (ed è unico; l'unicità è una banale conseguenza della ipotesi di indipendenza lineare dei $\nabla h_i(x^\star)$) un vettore $\lambda^\star = (\lambda_1^\star, \dots, \lambda_m^\star) \in \mathbb{R}^m$ tale che:
$\nabla f(x^\star) - \sum_{i=1}^{m} \lambda_i^\star \nabla h_i(x^\star) = 0$.
Naturalmente $x^\star$ soddisfa anche la condizione $h(x^\star) = 0$ (cioè: $h_i(x^\star) = 0$ per ogni i) .

NOTA "FUORI TESTO"
Notare che la condizione di indipendenza lineare implica $m \le n$. A dire il vero, i caso $m = n$ è "normalmente" poco interessante: ci aspettiamo che il sistema dei vincoli abbia un numero finito di soluzioni, ovvero di punti in $\mathbb{R}^n$ e quindi il problema di massimo diventa relativamente banale.
Si noti inoltre che, se $m = n$, tenendo conto della condizione di indipendenza lineare dei gradienti, la tesi del teorema di Lagrange è una banalità: è ovvio che esistano e siano unici $\lambda_i^\star$ tali che $\nabla f(x^\star) = \sum_{i \in U} \lambda_i^\star \nabla h_i(x^\star)$, visto che n vettori linearmente indipendenti in uno spazio vettoriale di dimensione n sono una base per tale spazio e quindi *ogni* vettore (non solo $\nabla f(x^\star)$) è esprimibile come loro combinazione lineare.
FINE NOTA "FUORI TESTO".

Possiamo introdurre la lagrangiana: $L(x, \lambda) = f(x) - \sum_{i=1}^{m} \lambda_i h_i(x)$.
Le condizioni le possiamo riscrivere allora così:.
$$\begin{cases} \nabla_x L(x^\star, \lambda^\star) = 0 \\ \nabla_\lambda L(x^\star, \lambda^\star) = 0 \end{cases}$$
Mi riferirò a questo sistema, se necessario, come al sistema (L).

ESEMPIO:
$f(x, y) = x$ con vincolo $x + y = 1$. Le condizioni del teorema sono soddisfatte (in particolare, l'unico vettore $\nabla h(x, y)$ è costante ed uguale a $(1, 1)$. Quindi non si annulla mai e pertanto neanche sul vincolo E. Pertanto l'insieme dei vettori $\nabla h(x, y)$ è sempre un insieme linearmente indipendente (per un insieme che contiene un solo vettore ciò equivale a verificare che questo vettore non sia il vettore nullo).
Non troviamo nessun punto che risolva il sistema (L).
Il sistema (L), infatti, è:
$$\begin{cases} 1 - \lambda = 0 \\ -\lambda = 0 \\ x - \lambda(x + y - 1) = 0 \end{cases}$$
E, chiaramente, le prime due equazioni sono contraddittorie fra loro.
D'altronde, è evidente "a occhio" che il problema dato non presenta alcun punto di max locale.

ESEMPIO:
$f(x, y) = x$ con vincolo $x^2 + y^2 = 1$. Le condizioni del teorema sono soddisfatte (in particolare, l'unico vettore $\nabla h(x, y)$ vale $(2x, 2y)$ e quindi non si annulla **sul vincolo**).
Il sistema (L) è:
$$\begin{cases} 1 - 2\lambda x = 0 \\ -2\lambda y = 0 \\ x^2 + y^2 = 1 \end{cases}$$
Questo sistema ha due soluzioni, le cui coordinate (x, y) sono, rispettivamente: $(1, 0)$ e $(-1, 0)$.
Usando il teorema di Weierstrass si vede che uno, il punto $(1, 0)$, è punto di max globale. L'altro è punto di min globale.

Interpretazione geometrica dei moltiplicatori di Lagrange (con due vincoli, in $\mathbb{R}^3$):
- il ∇f sta nello spazio vettoriale generato da ∇h_1 e ∇h_2
- il ∇f è ortogonale allo spazio vettoriale approssimazione lineare dei vincoli (Bertsekas lo chiama "sottospazio delle variazioni ammissibili del primo ordine")

La dimostrazione (meglio, una delle possibili dimostrazioni) utilizza il teorema delle funzioni implicite (teorema di Dini). Lo si può dimostrare anche usando una idea, che è usata per la soluzione numerica di problemi di minimo vincolato: il metodo di penalizzazione.

ESEMPIO:
Il significato geometrico del teorema dei moltiplicatori di Lagrange può essere visualizzato confrontando un paio di esempi. In entrambi i casi i vincoli sono: $h_1(x, y, z) = x^2 + y^2 + z^2 - 1 = 0$ e $h_2(x, y, z) = x = 0$. Le condizioni del teorema sono soddisfatte ($\nabla h_1(x, y, z) = (2x, 2y, 2z)$ e $\nabla h_2(x, y, z) = (1, 0, 0)$, che sono linearmente indipendenti sul vincolo).
Come funzione obiettivo si possono considerare due casi: $f(x, y, z) = x$ e $f(x, y, z) = x + y$.

ESEMPIO:
Cercare massimi e minimi globali di $(x + 2y)^2$ sul vincolo $\frac{x^2}{4} + \frac{y^2}{3} = 1$.
SOLUZIONE
Il vincolo è un sottoinsieme chiuso di $\mathbb{R}^2$, visto che è il luogo dei punti in cui si annulla la funzione, definita e continua su tutto $\mathbb{R}^2$, $(x, y) \mapsto \frac{x^2}{4} + \frac{y^2}{3}$. Questo luogo dei punti è anche, ovviamente, un sottoinsieme limitato di $\mathbb{R}^2$ (perché?), e quindi il teorema di Weierstrass ci garantisce che la nostra funzione obiettivo abbia massimo e minimo globali sul vincolo (che è un sottoinsieme chiuso, limitato e non vuoto di $\mathbb{R}^2$).
Visto che un punto di estremo (cioè, di max o di min) globale è anche un punto di estremo locale, e visto che sono soddisfatte le ipotesi del teorema di Lagrange (verificare!), i punti di estremo globale andranno cercati fra i punto che soddisfano le CN di Lagrange.

Abbiamo il sistema:

$$\begin{cases} 2(x+2y) - \lambda\dfrac{x}{2} = 0 \\ 4(x+2y) - \lambda\dfrac{2y}{3} = 0 \\ \dfrac{x^2}{4} + \dfrac{y^2}{3} = 1 \end{cases}$$

Dividiamo entrambi i membri della seconda equazione per 2 e sottraiamo la seconda equazione alla prima. Otteniamo la condizione $-\lambda\left(\frac{x}{2} - \frac{y}{3}\right) = 0$.

Questa equazione è soddisfatta se:
- $\lambda = 0$, e quindi il sistema si riduce a:

$$\begin{cases} 2(x+2y) = 0 \\ \dfrac{x^2}{4} + \dfrac{y^2}{3} = 1 \end{cases}$$

Vale a dire, i punti in cui la retta di equazione $x + 2y = 0$ incontra il vincolo (che è un'ellisse).
- $\frac{x}{2} - \frac{y}{3} = 0$. Si lascia al lettore il facile calcolo.

A questo punto, si tratta solo di calcolare il valore della funzione obiettivo nei punti trovati e selezionare quelli più alti (cui corrispondono punto di max) e quelli più bassi (punti di min).
Nota finale: per come è la funzione obiettivo, era ovvio che i punti di minimo (valore minimo è 0) fossero assunti sulla retta $(x + 2y) = 0$

ESEMPIO (importanza della condizione di indipendenza lineare):
Senza la condizione di lineare indipendenza, il teorema è falso. Prendere: $f(x, y, z) = x + y + z$ (ad esempio, ma c'è ampia scelta...) e i vincoli siano $h_1(x, y, z) = x^2 + y^2 + (z-1)^2 - 1 = 0$ e $h_1(x, y, z) = x^2 + y^2 + (z-2)^2 - 4 = 0$. Il vincolo descrive un singleton. Ma il sistema (L) non ha soluzione:

$$\begin{cases} 1 + 2\lambda x + 2\mu x = 0 \\ 1 + 2\lambda y + 2\mu y = 0 \\ 1 + 2\lambda(z-1) + 2\mu(z-2) = 0 \\ x^2 + y^2 + (z-1)^2 = 1 \\ x^2 + y^2 + (z-2)^2 = 4 \end{cases}$$

Infatti dalle due ultime equazioni ricaviamo che $(x, y, z) = (0, 0, 0)$. Sostituendo nella prima equazione troviamo $1 = 0$.
Tutto normale: non essendo soddisfatte le condizioni del teorema dei moltiplicatori di Lagrange non abbiamo nessuna garanzia che il sistema (L) abbia soluzione (anche se non possiamo escludere che ce l'abbia, in linea generale).

DOCUMENTI:
.
ESERCIZI:
.

NOTA:
Il metodo dei moltiplicatori di Lagrange ha una sua dimostrazione "naturale" fondata sull'uso del teorema delle funzioni implicite (teorema di Dini).
Ne ricordo brevemente l'enunciato, nel caso più semplice, ovvero con una condizione sola che riguarda una funzione di due variabili.
Consideriamo $f: A \to \mathbb{R}$, $A \subseteq \mathbb{R}$, A aperto, e con $f \in \mathcal{C}^1(A)$. Siamo interessati al luogo dei punti che soddisfano la condizione $f(x, y) = 0$.
Definiamo $E = \{(x, y) \in A : f(x, y) = 0\}$
Si può provare che, sotto opportune ipotesi, questo luogo geometrico (che è un "insieme di livello" per f) è, localmente, grafico di una funzione di una sola variabile.
Sia $(\bar{x}, \bar{y})$ t.c. $f(\bar{x}, \bar{y}) = 0$. Con $\frac{\partial f}{\partial y}(\bar{x}, \bar{y}) \neq 0$.
Allora esiste $\delta > 0$ t.c. in $I_\delta(\bar{x}, \bar{y}) = \left\{(x, y) \in \mathbb{R}^2 : ||(x, y) - (\bar{x}, \bar{y})|| < \delta\right\}$ lo insieme di livello è grafico di una funzione di una variabile.
Cioè, esiste una funzione φ, definita in un opportuno intorno $I =]\bar{x} - \alpha, \bar{x} + \beta[$ di $\bar{x}$ t.c. (ovviamente $\alpha, \beta > 0$) per ogni $(x, y) \in I_\delta(\bar{x}, \bar{y})$: $f(x, y) = 0$ se e solo se $y = \varphi(x)$.
Non solo: questa funzione φ è derivabile, con derivata prima continua, e si ha (per ogni $x \in I$):
$\varphi'(x) = -\dfrac{f_x(x, \varphi(x))}{f_y(x, \varphi(x))}$.

E' facile, avendo a disposizione questo risultato, provare che la regolarità di f si riflette direttamente nella regolarità di φ. In generale, se f è di classe $\mathcal{C}^k(A)$, allora anche φ è di classe $\mathcal{C}^k(A)$. Vediamo, ad esempio, come si prova questo fatto per $k = 2$.
Se f è di classe $\mathcal{C}^2(A)$, allora f_x ed f_y sono di classe $\mathcal{C}^1(A)$. Quindi, la funzione $x \mapsto -\dfrac{f_x(x, \varphi(x))}{f_y(x, \varphi(x))}$ è una funzione di classe $\mathcal{C}^1(A)$ in quanto frazione di funzioni che sono di classe $\mathcal{C}^1(A)$ in quanto composte di funzioni di classe $\mathcal{C}^1(A)$. Ciò vale su I, intervallo aperto sul quale questa funzione coincide con $\varphi'(x)$. Allora, $\varphi'(x)$ è di classe $\mathcal{C}^1(A)$ sull'intervallo aperto I in quanto coincide, su questo intervallo, con una funzione di classe $\mathcal{C}^1(A)$. Allora, se φ' è di classe $\mathcal{C}^1(A)$, φ è di classe $\mathcal{C}^2(A)$, come affermato.

Un esempio di applicazione del teorema di Dini riguarda la dipendenza delle radici di un polinomio dai coefficienti. Lo vediamo in un caso particolare, di rilevanza applicativa: il calcolo del TIR (Tasso Interno di Rendimento; in inglese IRR, ovvero Internal Rate of Return).
Un progetto di investimento è una (n+1)-pla $(a_0, a_1, \ldots, a_n)$ (gli elementi di questo vettore rappresentano i flussi di cassa: a_k indica la quantità

di soldi, entranti se positivo, sennò uscenti). Nel nostro caso ci occuperemo di un caso particolare, che descrive un *progetto di finanziamento* (o di "provvista"), nel senso che supporremo: $a_0 > 0$ e $a_k < 0$ per $k \in \{1, ..., n\}$. Stiamo immaginando di ottenere un prestito a_0 che poi dobbiamo restituire con una qualche forma di rateizzazione.
Facciamo questo caso particolare perché con le ipotesi date è facile provare (classico esercizio di Analisi I) che esiste ed è unico v t.c. $F(a_0, a_1, ..., a_n, v) = a_0 + a_1 vu + ... + a_n v^n = 0$. Assumiamo, per semplicità, che i vari flussi di cassa abbiano uno scadenziario fisso, diciamo annuale. La variabile v indica il tasso di sconto: ricordo che $v = \frac{1}{1+i}$, dove i rappresenta il tasso di interesse. Il TIR è dato dallo i corrispondente al v che risolve questa equazione algebrica.

Bene, dati $\bar{a}_0, \bar{a}_1, ..., \bar{a}_n$ (soddisfacenti le condizioni date sopra sul loro segno), sia $\bar{v}$ l'unica soluzione di $F(\bar{a}_0, \bar{a}_1, ..., \bar{a}_n, v) = 0$. La domanda che ci poniamo (per semplicità) è come varia v al variate del coefficiente $\bar{a}_1$, tenendo costanti gli altri coefficienti.
Definiamo $f(a_1, v) = F(\bar{a}_0, a_1, \bar{a}_2, ..., \bar{a}_n, v)$.
E' $f(\bar{a}_1, \bar{v}) = 0$. Ovviamente f è di classe $C^1(A)$, con $A = \mathbb{R}^2$. Vediamo se è soddisfatta la condizione su $\frac{\partial f}{\partial v}(\bar{a}_1, v) =\neq 0$. La risposta è molto semplice: questa derivata parziale ci dà un polinomio che ha tutti i coefficienti strettamente negativi e $v > 0$, quindi ovviamente è diversa da zero.
Pertanto abbiamo a disposizione la funzione $v = \varphi(a_1)$. Volendo, possiamo calcolare $\varphi'(a_1)$, grazie al teorema di Dini. Il che ci permette di dire quale sia la "sensitività" di v a variazioni del parametro $\bar{a}_1$.

ESEMPIO:
Vediamo un caso semplicissimo, in modo da poter fare i conti senza usare il teorema di Dini, e quindi verificare se e quanto le conclusioni che possiamo trarre con Dini collimano con i conti esatti fatti nell'esempio. Supponiamo che la restituzione avvenga mediante il pagamento di una "lump sum" ad una data successiva (ovviamente?) a quella in cui si è ricevuto il prestito. Questa data la contrassegniamo come periodo 1 (per comodità possiamo pensare che la restituzione avvenga a un anno dal prestito, così possiamo parlare tranquillamente di tassi di interesse e di sconto annui).
Abbiamo allora: $(\bar{a}_0, \bar{a}_1)$. Il v che risolve l'equazione $\bar{a}_0 + \bar{a}_1 v = 0$ è $-\frac{\bar{a}_0}{\bar{a}_1}$. Se teniamo $\bar{a}_0$ fisso, abbiamo che $\varphi(a_1) = -\frac{\bar{a}_0}{a_1}$ la cui derivata (rispetto ad a_1, naturalmente) è $\varphi'(a_1) = \frac{\bar{a}_0}{a_1^2}$.
Supponiamo che $\bar{a}_0 = 100$ e $\bar{a}_1 = -110$. Ovviamente $i = 0.1$ (ovvero, 10% annuo), e altrettanto ovviamente $v = \frac{100}{110} \sim 0.909090909$.
L'uso del teorema di Dini ci permettere di calcolare la derivata di φ nel punto a_1 (sto assumendo che siamo un po' stupidini, ovvero che non usiamo la forma esplicita di φ che in questo caso semplicissimo abbiamo già determinato: ma questo "far finta di niente" è per vedere come funziona il teorema di Dini laddove non fossimo capaci di trovare φ). Vediamo cosa otteniamo.
Immaginiamo una piccola variazione di a_1, ovvero che diventi 111 invece di 110. Di quanto varia v? Facendo i conti espliciti, otteniamo $v = \frac{100}{111} \sim 0.900900901$ (ovviamente $i = 0.11$).
Con Dini, usando $\varphi'(110) = -\frac{100}{12100}$, abbiamo che
$\varphi(111) \sim \varphi(110) + \varphi'(110) \cdot (111 - 110) \sim 0,909090909 + 0,008264463 \cdot 1 = 0,900826446$. Come si vede, una approssimazione ragionevole: la differenza tra il valore vero e quello approssimato è pari a $0,000074455$. Stiamo parlando di v, ma i discorsi si trasferiscono molto semplicemente ad i, ovvero al TIR.
LINK:
Segnalo, per chi fosse interessato, questi appunti in rete. E' descritto il teorema di Dini, il teorema della funzione implicita, l'applicazione del teorema di Dini alla ricerca di estremi vincolati.

Ma 17 novembre 2008, 2h-10h:
Se abbiamo anche vincoli di disuguaglianza, ci sono le condizioni di Kuhn-Tucker.
Stavolta il vincolo è dato da $h_i(x) = 0$ e $g_j(x) \leq 0$ (e assumiamo D non vuoto)
Serve l'idea di vincolo *attivo*. Un vincolo di disuguaglianza (il vincolo j) è attivo in un punto $x^\star$ se $g_j(x^\star) = 0$. Indichiamo con $A(x^\star)$ l'insieme degli *indici* [nota bene: $A(x^\star) \subseteq D$] che identificano i vincoli attivi nel punto $x^\star$
Allora, se (oltre alle altre ovvie condizioni) la famiglia dei vettori:
- $(\nabla h_i(x^\star))_{i \in U}$
- $\nabla g_j(x^\star)$ con $j \in A(x^\star)$
è una famiglia di vettori linearmente indipendenti,
esiste ed è unico $(\lambda^\star, \mu^\star) = \left((\lambda_i^\star)_{i \in U}, (\mu_j^\star)_{j \in D}\right)$ t.c. (introducendo la lagrangiana: $L(x, \lambda, \mu) = f(x) - \sum_{i \in U} \lambda_i h_i(x) - \sum_{j \in D} \mu_j g_j(x)$):

$$\begin{cases} \nabla_x L(x^\star, \lambda^\star, \mu^\star) = 0 \\ \nabla_\lambda L(x^\star, \lambda^\star, \mu^\star) = 0 \\ \nabla_\mu L(x^\star, \lambda^\star, \mu^\star) \geq 0 \\ \mu^\star \geq 0 \\ \mu^\star \cdot \nabla_\mu L(x^\star, \lambda^\star, \mu^\star) = 0 \end{cases}$$

Chiamerò "sistema (KT)" questo sistema di equazioni e disequazioni.

La seconda e terza equazione nel sistema non sono altro che le condizioni dei vincoli. Possiamo allora riscrivere il sistema così

$$\begin{cases} \nabla_x L(x^\star, \lambda^\star, \mu^\star) = 0 \\ h(x^\star) = 0 \\ g(x^\star) \le 0 \\ \mu^\star \ge 0 \\ \mu^\star \cdot g(x^\star) = 0 \end{cases}$$

L'ultima condizione è $\sum_{j \in D} \mu_j^\star \cdot g_j(x^\star) = 0$. Ma, viste le condizioni sui segni di $\mu_j^\star$ e $g_j(x^\star)$, *ogni addendo* di questa sommatoria è minore o uguale a zero. Grazie a questa osservazione possiamo concludere che ogni addendo è uguale a 0. Ovvero: $\mu_j^\star \cdot g_j(x^\star) = 0$ per ogni j

Una osservazione terminologica. Le ultime tre condizioni del sistema (KT), nel loro complesso, vengono chiamate condizioni di complementarità. Il sistema seguente è un problema di complementarità (non lineare, visto che non si fa una ipotesi di linearità sulla funzione g):

$$\begin{cases} g(x) \le 0 \\ \mu \ge 0 \\ \mu \cdot g(x) = 0 \end{cases}$$

ESEMPIO:
$f(x, y) = x$ con vincolo $x^2 + y^2 \le 1$. Vediamo cosa succede per $(x, y) = (1, 0)$, che è chiaramente punto di max assoluto per f col vincolo dato. Il vincolo è attivo in $(1, 0)$. Le condizioni del teorema sono soddisfatte (in particolare, l'unico vettore $\nabla g(x, y)$ vale $(2x, 2y)$ e quindi non si annulla in $(1, 0)$).
Il sistema (KT) è:

$$\begin{cases} 1 - 2\mu x = 0 \\ -2\mu y = 0 \\ x^2 + y^2 - 1 \le 0 \\ \mu \ge 0 \\ \mu\left(x^2 + y^2 - 1\right) = 0 \end{cases}$$

Visto che il vincolo è attivo (infatti nel punto $(1, 0)$ si ha $x^2 + y^2 - 1 = 0$), l'ultima condizione è superflua e il sistema (KT) diviene:

$$\begin{cases} 1 - 2\mu x = 0 \\ -2\mu y = 0 \\ x^2 + y^2 = 1 \\ \mu \ge 0 \end{cases}$$

Ovviamente $\left(1, 0, \frac{1}{2}\right)$ è soluzione del sistema. Notare che il moltiplicatore vale $\frac{1}{2}$ (per K-T doveva essere maggiore o uguale a zero).

ESEMPIO:
$f(x, y) = x$ con vincolo $x^2 + y^2 \le 1$ e $x + y \le 5$. Vediamo cosa succede per $(x, y) = (1, 0)$, che è chiaramente punto di max assoluto per f col vincolo dato. Il primo vincolo è attivo in $(1, 0)$, il secondo no. Le condizioni del teorema sono soddisfatte (in particolare, l'unico vettore $\nabla g_1(x, y)$ che ci interessa, in quanto corrispondente all'unico vincolo attivo, vale $(2x, 2y)$ e quindi non si annulla in $(1, 0)$).
Il sistema (KT) è:

$$\begin{cases} 1 - 2\mu_1 x - \mu_2 = 0 \\ -2\mu y - \mu_2 = 0 \\ x^2 + y^2 - 1 \le 0 \\ x + y - 5 \le 0 \\ \mu_1 \ge 0 \\ \mu_2 \ge 0 \\ \mu_1\left(x^2 + y^2 - 1\right) = 0 \\ \mu_2(x + y - 5) = 0 \end{cases}$$

Visto che g_1 è attivo come vincolo, e g_2 no, il sistema (KT) diventa:

$$\begin{cases} 1 - 2\mu_1 x - \mu_2 = 0 \\ -2\mu y - \mu_2 = 0 \\ x^2 + y^2 = 1 \\ x + y < 5 \\ \mu_1 \ge 0 \\ \mu_2 = 0 \end{cases}$$

Ovviamente $\left(1, 0, \frac{1}{2}, 0\right)$ è soluzione del sistema. Notare che il primo moltiplicatore vale $\frac{1}{2}$ (per K-T doveva essere maggiore o uguale a zero), mentre il secondo vale 0.

OSSERVAZIONE:

Specificare che l'indipendenza lineare dei gradienti delle g_j è richiesta solo per i vincoli **attivi** è essenziale. E' del tutto normale la presenza di un grande numero di vincoli di disuguaglianza, maggiore di n: se venisse richiesta l'indipendenza lineare di tutti i ∇g_j, essa non potrebbe aversi (il massimo numero di vettori linearmente indipendenti in $\mathbb{R}^n$ è n).
Si può fare un semplicissimo esempio, in $\mathbb{R}^2$. Si considerino i vincoli:
$x \geq 0$, $y \geq 0$ e $x + y \leq 1$.
Ovvero: $g_1(x, y) = -x \leq 0$, $g_2(x, y) = -y \leq 0$ e $g_3(x, y) = x + y - 1 \leq 0$.
I gradienti di queste tre funzioni sono: $(-1, 0)$, $(0, -1)$ e $(1, 1)$, che ovviamente sono linearmente dipendenti.
E' immediato rendersi conto, tuttavia, che in ogni punto che soddisfa i vincoli, il massimo numero di vincoli attivi che si hanno è pari a 2.

OSSERVAZIONE:
Sia dato un punto che risolve il sistema offerto dalle condizioni di Kuhn-Tucker. Cioè:

$$\begin{cases} \nabla_x L(x^\star, \lambda^\star, \mu^\star) = 0 \\ h(x^\star) = 0 \\ g(x^\star) \leq 0 \\ \mu^\star \geq 0 \\ \mu^\star \cdot g(x^\star) = 0 \end{cases}$$

Supponiamo che in quel punto non ci sia nessun vincolo attivo. Il sistema ci dice che $\mu^\star = 0$ (immediato dall'ultima condizione, visto che $g(x^\star) > 0$). Quindi, abbiamo:

$$\begin{cases} \nabla_x L(x^\star, \lambda^\star, \mu^\star) = 0 \\ h(x^\star) = 0 \\ g(x^\star) < 0 \\ \mu^\star = 0 \end{cases}$$

Insomma, sapendo che $g(x^\star) < 0$, il sistema si riduce al sistema ottenuto dai moltiplicatori di Lagrange (a parte la inutile informazione che $\mu^\star = 0$), come se sopprimessimo i vincoli di disuguaglianza.

Ve 20 novembre 2008, 3h-13h:
Naturalmente le CN di K-T servono a **trovare** punti "sospettati" di essere di massimo locale. Quindi, non abbiamo a disposizione il punto $\bar{x}$, come nei due esempi illustrativi precedenti. Il problema è quello di trovare $\bar{x}$. E la ricerca si effettua scrivendo le condizioni di K-T e cercando i punti (e moltiplicatori) che le soddisfano. Di fatto, ci troviamo a risolvere un sistema di equazioni e disequazioni.

ESEMPIO SVOLTO:
$\max\ x + 2y + 3z$ con vincoli: $x^2 + y^2 + z^2 \leq 4$ e $z \geq 1$.
Analizzati i 4 diversi casi che si possono presentare, a seconda di chi siano i vincoli attivi.

Osserviamo che sia la funzione f data ($f(x, y, z) = x + 2y + 3z$) che le funzioni g_1 e g_2 che descrivono i vicoli sono definite su $\mathbb{R}^2$ e sono anche di classe $C^1\left(\mathbb{R}^2\right)$.

Per quanto riguarda la lineare indipendenza dei gradienti dei vincoli, notiamo che:
$g_1(x, y, z) = x^2 + y^2 + z^2 - 4$ e $g_2(x, y, z) = 1 - z$ (per g_2 occorre fare attenzione che il vincolo era stato dato in forma di ≥ 0 invece che di ≤ 0).
Quindi $\nabla g_1(x, y, z) = (2x, 2y, 2z)$ e $\nabla g_2(x, y, z) = (0, 0, -1)$.

I due gradienti sono linearmente dipendenti se e solo se $x = y = 0$. Ma queste condizioni ci dicono che siamo sull'asse delle z. Ora, sull'asse delle zeta, troviamo solo due punti che appartengono al vincolo: $(0, 0, 1)$ e $(0, 0, 2)$. Nel primo punto il solo vincolo attivo è g_2 e nel secondo punto è solo g_1. Ma in questi punti il gradiente di g_2 (rispettivamente, di g_1) non si annulla. Pertanto la condizione di non degenerazione è soddisfatta.

Allora le condizioni di KT sono necessarie. Scriviamo il sistema che ci dà le condizioni di KT. La lagrangiana è:
$L\left(x, y, z, \mu_1, \mu_2\right) = f(x, y, z) - \mu_1 g_1(x, y, z) - \mu_2 g_2(x, y, z) = x + 2y + 3z + \mu_1\left(x^2 + y^2 + z^2 - 4\right) - \mu_2(1 - z)$:

$$\begin{cases} \nabla_x L\left(x, y, z, \mu_1, \mu_2\right) = 0 \\ \dfrac{\partial L\left(x, y, z, \mu_1, \mu_2\right)}{\partial \mu_1} \leq 0 \\ \dfrac{\partial L\left(x, y, z, \mu_1, \mu_2\right)}{\partial \mu_2} \leq 0 \\ \mu_1 \geq 0 \\ \mu_2 \geq 0 \\ \mu_1 \dfrac{\partial L\left(x, y, z, \mu_1, \mu_2\right)}{\partial \mu_1} = 0 \\ \mu_2 \dfrac{\partial L\left(x, y, z, \mu_1, \mu_2\right)}{\partial \mu_2} = 0 \end{cases}$$

Chiamo *disuguaglianze* (D) il seguente sistema:

$$\begin{cases} \dfrac{\partial L(x,y,z,\mu_1,\mu_2)}{\partial \mu_1} \le 0 \\ \dfrac{\partial L(x,y,z,\mu_1,\mu_2)}{\partial \mu_2} \le 0 \\ \mu_1 \ge 0 \\ \mu_2 \ge 0 \end{cases}$$

Usando la legge dell'annullamento del prodotto, possiamo dire che le soluzioni del sistema di KT sono date dalle soluzioni di almeno uno dei 4 sistemi seguenti, che indichiamo con (1), (2), (3) e (4).

Sistema (1):

$$\begin{cases} \nabla_x L(x,y,z,\mu_1,\mu_2) = 0 \\ \mu_1 = 0 \\ \mu_2 = 0 \\ \textit{disuguaglianze } (D) \end{cases}$$

Sistema (2):

$$\begin{cases} \nabla_x L(x,y,z,\mu_1,\mu_2) = 0 \\ \mu_1 = 0 \\ \dfrac{\partial L(x,y,z,\mu_1,\mu_2)}{\partial \mu_2} = 0 \\ \textit{disuguaglianze } (D) \end{cases}$$

Sistema (3):

$$\begin{cases} \nabla_x L(x,y,z,\mu_1,\mu_2) = 0 \\ \dfrac{\partial L(x,y,z,\mu_1,\mu_2)}{\partial \mu_1} = 0 \\ \mu_2 = 0 \\ \textit{disuguaglianze } (D) \end{cases}$$

Sistema (4):

$$\begin{cases} \nabla_x L(x,y,z,\mu_1,\mu_2) = 0 \\ \dfrac{\partial L(x,y,z,\mu_1,\mu_2)}{\partial \mu_1} = 0 \\ \dfrac{\partial L(x,y,z,\mu_1,\mu_2)}{\partial \mu_2} = 0 \\ \textit{disuguaglianze } (D) \end{cases}$$

Riscriviamoli in dettaglio nel nostro caso:

Sistema (1):

$$\begin{cases} 1 - 2\mu_1 x = 0 \\ 2 - 2\mu_1 y = 0 \\ 3 - 2\mu_1 z + \mu_2 = 0 \\ \mu_1 = 0 \\ \mu_2 = 0 \\ \textit{disuguaglianze } (D) \end{cases} \quad \text{ovvero} \quad \begin{cases} 1 = 0 \\ 2 = 0 \\ 3 = 0 \\ \mu_1 = 0 \\ \mu_2 = 0 \\ \textit{disuguaglianze } (D) \end{cases}$$

Ovviamente non ha soluzioni (nessun punto "interno al vincolo" soddisfa le CN).

Sistema (2):

$$\begin{cases} 1 - 2\mu_1 x = 0 \\ 2 - 2\mu_1 y = 0 \\ 3 - 2\mu_1 z + \mu_2 = 0 \\ \mu_1 = 0 \\ 1 - z = 0 \\ \textit{disuguaglianze } (D) \end{cases} \quad \text{ovvero} \quad \begin{cases} 1 - = 0 \\ 2 - = 0 \\ 3 + \mu_2 = 0 \\ \mu_1 = 0 \\ 1 - z = 0 \\ \textit{disuguaglianze } (D) \end{cases}$$

Anche per questo, ovviamente, nessuna soluzione (nessun punto sul piano di equazione $1 - z = 0$, e che appartenga al vincolo, soddisfa le CN).

Sistema (3):

$$\begin{cases} 1-2\mu_1 x=0 \\ 2-2\mu_1 y=0 \\ 3-2\mu_1 z+\mu_2=0 \\ x^2+y^2+z^2-4=0 \\ \mu_2=0 \\ \textit{disuguaglianze } (D) \end{cases} \quad \text{ovvero} \quad \begin{cases} x=\dfrac{1}{2\mu_1} \\ y=\dfrac{2}{2\mu_1} \\ z=\dfrac{3}{2\mu_1} \\ \dfrac{1}{4\mu_1^2}+\dfrac{4}{4\mu_1^2}+\dfrac{9}{4\mu_1^2}=4 \\ \mu_2=0 \\ \textit{disuguaglianze } (D) \end{cases} \quad \text{ovvero} \quad \begin{cases} x=\dfrac{1}{2\mu_1} \\ y=\dfrac{2}{2\mu_1} \\ z=\dfrac{3}{2\mu_1} \\ \mu_1=\pm\sqrt{\dfrac{7}{8}} \\ \mu_2=0 \\ \textit{disuguaglianze } (D) \end{cases} \quad \text{ovvero} \quad \begin{cases} x=\dfrac{1}{2}\sqrt{\dfrac{8}{7}} \\ y=\dfrac{2}{2}\sqrt{\dfrac{8}{7}} \\ z=\dfrac{3}{2}\sqrt{\dfrac{8}{7}} \\ \mu_1=\sqrt{\dfrac{7}{8}} \\ \mu_2=0 \\ \textit{disuguaglianze } (D) \end{cases}$$

Nell'ultimo passaggio abbiamo usato le $\textit{disuguaglianze } (D)$ per eliminare la soluzione corrispondente a $\mu_1=-\sqrt{\frac{7}{8}}$. L'unica soluzione è quindi: $\left(\frac{1}{2}\sqrt{\frac{8}{7}}, \sqrt{\frac{8}{7}}, \frac{3}{2}\sqrt{\frac{8}{7}}, \sqrt{\frac{7}{8}}, 0\right)$.

Sistema (4):

$$\begin{cases} 1-2\mu_1 x=0 \\ 2-2\mu_1 y=0 \\ 3-2\mu_1 z+\mu_2=0 \\ x^2+y^2+z^2-4=0 \\ 1-z=0 \\ \textit{disuguaglianze } (D) \end{cases} \quad \text{ovvero} \quad \begin{cases} x=\dfrac{1}{2\mu_1} \\ y=\dfrac{2}{2\mu_1} \\ 3-2\mu_1+\mu_2=0 \\ x^2+y^2=3 \\ z=1 \\ \textit{disuguaglianze } (D) \end{cases} \quad \text{ovvero} \quad \begin{cases} x=\dfrac{1}{2\mu_1} \\ y=\dfrac{2}{2\mu_1} \\ 3-2\mu_1+\mu_2=0 \\ \dfrac{1}{4\mu_1^2}+\dfrac{4}{4\mu_1^2}=3 \\ z=1 \\ \textit{disuguaglianze } (D) \end{cases} \quad \text{ovvero} \quad \begin{cases} x=\dfrac{1}{2\mu_1} \\ y=\dfrac{2}{2\mu_1} \\ 3-2\mu_1+\mu_2=0 \\ \mu_1=\pm\sqrt{\dfrac{5}{12}} \\ z=1 \\ \textit{disuguaglianze } (D) \end{cases}$$

$$\text{ovvero} \quad \begin{cases} x=\dfrac{1}{2\mu_1} \\ y=\dfrac{2}{2\mu_1} \\ 3-2\mu_1+\mu_2=0 \\ \mu_1=\sqrt{\dfrac{5}{12}} \\ z=1 \\ \textit{disuguaglianze } (D) \end{cases} \quad \text{ovvero} \quad \begin{cases} x=\dfrac{1}{2\mu_1} \\ y=\dfrac{2}{2\mu_1} \\ \mu_2=2\sqrt{\dfrac{5}{12}}-3 \\ \mu_1=\sqrt{\dfrac{5}{12}} \\ z=1 \\ \textit{disuguaglianze } (D) \end{cases}$$

Si perviene al penultimo sistema grazie alle $\textit{disuguaglianze } (D)$ che escludono la soluzione corrispondente a $\mu_1=-\sqrt{\frac{5}{12}}$. Nell'ultimo sistema è evidente che μ_2 non è ≥ 0, e quindi concludiamo che il sistema (4) non ha soluzioni.

Il sistema (4) mostra un fenomeno interessante. Rende evidente che quanto si fa non corrisponde alla ricerca di punti che soddisfino la CN di max locale ristretti ai vincoli: $x^2+y^2+z^2-4=0$ e $1-z=0$. In questo caso, vi sarebbe per forza un punto di max globale (Weierstrass). Mentre il sistema (4) non ci fornisce nessuna soluzione. L'apparente "mistero" è facilmente risolubile. Il sistema (4) cerca le soluzioni **del pb dato** che **si trovino sul vincolo** $x^2+y^2+z^2-4=0$ e $1-z=0$. **Non** cerca i punti di max di f **ristretta** al vincolo $x^2+y^2+z^2-4=0$ e $1-z=0$.
La differenza si vede benissimo nel seguente esempio:
$\max z$ con vincoli: $x^2+y^2-z\leq 0$ e $1-z\leq 0$ (cioè, i punti dello spazio che stanno sia sopra al paraboloide di sezione circolare $z=x^2+y^2$ che al piano di equazione $z=1$).
Ovviamente tutti i punti della circonferenza (C) individuata (nello spazio) da $z=x^2+y^2$ e $z=1$ sono punti di massimo per f ristretta a tale vincolo (la f è costante e vale 1 sulla circonferenza (C)). Mentre, se scriviamo l'equivalente del sistema (4) in questo caso, vediamo che non ha soluzioni, come mostrato sotto. E questo è corretto, perché se ci si sposta verso l'alto, da un qualunque punto della circonferenza (C), il valore della funzione obiettivo aumenta e si resta dentro al vincolo dato. Quindi, il problema dato ovviamente non ha punti di max locale sulla circonferenza (C).
Sistema (4):

$$\begin{cases} 2\mu_1 x=0 \\ 2\mu_1 y=0 \\ 1+\mu_1+\mu_2=0 \\ x^2+y^2-z=0 \\ 1-z=0 \\ \textit{disuguaglianze } (D) \end{cases} \quad \text{ovvero} \quad \left[\begin{cases} \mu_1\neq 0 \\ x=0 \\ y=0 \\ 1+\mu_1+\mu_2=0 \\ z=0 \\ z=1 \\ \textit{disuguaglianze } (D) \end{cases} \quad \textit{vel} \quad \begin{cases} \mu_1=0 \\ 0=0 \\ 0=0 \\ \mu_2=-1 \\ 1=x^2+y^2 \\ z=1 \\ \textit{disuguaglianze } (D) \end{cases} \right]$$

Come si vede, nessuno dei due sistemi dentro la parentesi quadra ha soluzioni. In particolare, il secondo avrebbe come soluzione tutti i punti della circonferenza (C) se non ci fosse la condizione di non negatività sul moltiplicatore μ_2, che ci sta a ricordare che abbiamo "libertà di movimento" verso l'alto!

ESERCIZI:

.

Interpretazione geometrica:
Distinzione fra Lagrange e K-T:
-- Lagrange: $\nabla_x f = \lambda_1 \nabla_x h_1 + \lambda_2 \nabla_x h_2$. Il gradiente di f è nello spazio vettoriale generato da $\nabla_x h_1$ e $\nabla_x h_2$. (O, anche, il gradiente di f è ortogonale allo spazio vettoriale che ci dà la approssimazione lineare [locale] dei vincoli).
---- esempio (con due vincoli, in $\mathbb{R}^3$): $f(x, y, z) = ax + by + cz$, $h_1(x, y, z) = x$, $h_2(x, y, z) = y$: pertanto il vincolo è $x = 0, y = 0$, e quindi è l'asse delle z. Quindi $\nabla f(x, y, z) = (a, b, c)$. Solo se $c = 0$, il gradiente di f sta nel piano xy, che è lo spazio vettoriale generato da $\nabla h_1 = (1, 0, 0)$ e da $\nabla h_2 = (0, 1, 0)$.

-- K-T: $\nabla_x f = \mu_1 \nabla_x g_1 + \mu_2 \nabla_x g_2$. Con $\mu_1, \mu_2 \geq 0$. Il gradiente di f è nel **cono** generato da $\nabla_x g_1$ e $\nabla_x g_2$.
---- esempio (con due vincoli, in $\mathbb{R}^3$): $f(x, y, z) = ax + by + cz$, $g_1(x, y, z) = x$, $g_2(x, y, z) = y$. Pertanto il vincolo è $x \leq 0, y \leq 0$, cioè "l'angolo diedro" individuato, nel piano xy, dal terzo quadrante. Quindi $\nabla f(x, y, z) = (a, b, c)$. Solo se $c = 0$, il gradiente di f sta nel **primo quadrante** del piano xy, che è il cono generato da $\nabla g_1 = (1, 0, 0)$ e da $\nabla g_2 = (0, 1, 0)$.

Ma 24 novembre 2009, 2h-15h:
Algoritmi. Macchina di Turing. Per la definizione, vedi Hopcroft and Ullman, *Introduction to Automata Theory, Languages and Computation*, Addison-Wesley, Reading, 1979; un ottimo libro in merito è Marvin Minsky: *Computation: Finite and Infinite Machines*, Prentice-Hall, New Jersey (USA), 1967.
E' particolarmente rilevante la "tesi di Church" (o di "Church-Turing"), basata sulla dimostrazione che strade diverse usate per definire quali possano essere un "algoritmo" risultano essere equivalenti: le funzioni effettivamente calcolabili mediante ricorsione; mediante il λ-calcolo; mediante una macchina di Turing.
Criteri d'arresto (e loro problemi).
Esempi di criteri di arresto usati:
$\left\| x_{k+1} - x_k \right\| < \varepsilon$ (non garantisce convergenza: vedi serie che soddisfano la c.n. di convergenza ma non convegono, come $\sum_{i=1}^{\infty} \frac{1}{n}$
$\left| f(x_{k+1}) - f(x_k) \right| < \varepsilon$ (non garantisce convergenza: basta pensare a $f\colon [0, +\infty[$ definita così: $f(x) = x$ per $0 \leq x \leq 1$, e $f(x) = e^{-x}$. La successione k non converge, ma rende piccola a piacere la differenza usata come criterio di arresto).
$\left\| \nabla f(x_k) \right\| < \varepsilon$. Lascio al lettore trovare un esempio...
Metodo del gradiente.

DOCUMENTI:
Vedi i file: introduzione [da: http://www.dis.uniroma1.it/~or/gestionale/ottimizzazione/materiale_did/introduzione.pdf] e algoritmi [da: http://www.dis.uniroma1.it/~or/gestionale/ottimizzazione/materiale_did/algoritmi.pdf] dal sito [http://www.dis.uniroma1.it/~or/gestionale/ottimizzazione/] del corso di laurea in Ing Gest, Fac. di Ing., Univ. di Roma "La Sapienza".
Vedi anche il libro di Maffioli (per dettagli, vedi bibliografia [http://www.diptem.unige.it/patrone/Biblio_commentata_intro_RO.htm]).
Vedi anche le note di Sciandrone indicate in fondo a questo documento.
ESERCIZI:
.

Ma 27 novembre 2009
Non fatta.

Ma 1 dicembre 2009, 2h-17h:
Ordine di convergenza:
Esiste $c \in]0, 1[$ ed esiste $k_0 \in \mathbb{N}$ t.c. $\left\| x^{k+1} - x^{\star} \right\| \leq c \left\| x^k - x^{\star} \right\|$ per ogni $k \geq k_0$.

$p = 1$: ordine di convergenza lineare.

$1 < p < 2$: ordine di convergenza superlineare. (A volte si parla di ordine superlineare se $\lim_{k \to \infty} \frac{\left\| x^{k+1} - x^{\star} \right\|}{\left\| x^k - x^{\star} \right\|} = 0$).

$p = 2$: ordine di convergenza quadratico.

Metodo di Newton.

Programmazione lineare. In programmazione matematica avevamo $A \subseteq \mathbb{R}^n$ ed f, h_i, con $i \in U$ e g_j, $j \in D$. Tutte definite su A e a valori in $\mathbb{R}$
Cercavamo p.ti di max (locale, globale) per $f|\ E$ dove E è l'insieme dei punti che soddisfano il vincolo (o, brevemente, il vincolo). Cioè è l'insieme dei punti $x \in \mathbb{R}^n$ t.c. $h_i(x) = 0$ e $g_j(x) \leq 0$ per ogni $i \in U$ e $j \in D$.
Cosa cambia? Intanto, $A = \mathbb{R}^n$. E $f(x)$ è una funzione affine (ma possiamo assumerla lineare). Le funzioni h_i e g_j sono anch'esse affini.

Un esempio che sfrutteremo fino all'osso.

$\max x_1 + x_2$
con i vincoli:
$6x_1 + 4x_2 \leq 24$
$3x_1 - 2x_2 \leq 6$
$x_1, x_2 \geq 0$

Notiamo che l'insieme dei vincoli risulta essere chiuso e convesso. Anzi, è un *poliedro* convesso (intersezione di un numero *finito* di semispazi).
Ma abbiamo da massimizzare una funzione concava su un insieme convesso. Allora i p.ti di max locale sono anche di max globale.

Ecco perché ci interessiamo dei max globali e solo di quelli: occuparsi di max locali sarebbe lo stesso...
Abbiamo detto che abbiamo un poliedro convesso. L'intuizione ci porta a ritenere che, se c'è massimo, questo sarà assunto in almeno un vertice. Occorre stare attenti: un poliedro potrebbe non avere vertici: striscia di piano fra due rette parallele. Ma la condizione che imporremo (tutte le variabili non-negative) ci consentirà di poterlo affermare.

Allora si potrebbe pensare di trovare tutti i vertici e tra questi selezionare quello (o quelli) dove la funzione obiettivo assume il valore massimo. Ma ci sono metodi migliori (da molti punti di vista).
Tra questi, il metodo del simplesso. Si parte da un vertice e ci si sposta da quel vertice verso un vertice "adiacente" in modo da aumentare il valore della funzione obiettivo (vi saranno dei criteri che ci dicono se si può, e se la funzione può essere aumentata a piacere, essendo superiormente illimitata).

Visti 2 esempi.
- problema della dieta
- modello di trasporto

DOCUMENTI:
Vedi Tadei-Della Croce per esempi di modellizzazione mediante problemi di programmazione lineare (per dettagli, vedi bibliografia [http://www.diptem.unige.it/patrone/Biblio_commentata_intro_RO.htm]).

Ve 4 dicembre 2009, 2h-19h:
Una caratteristica specifica dei problemi di programmazione lineare è che, *se la funzione è superiormente limitata sul vincolo, allora ha massimo*. Questo non succede necessariamente per problemi di massimo non lineari: $-\exp(x)$, oppure $\mathrm{arctg}(x)$ sono esempi in tal senso.
La validità di questo fatto (superiormente limitato implica ha massimo) può essere provata, ad esempio, grazie alla decomposizione di 1kin, che vedremo in seguito.

Ogni problema di PL può essere messo in forma canonica. (Visto come trattare le variabili senza vincolo di segno: ad x sostituisco $u - v$, con $u, v \geq 0$. Un po' la stessa idea la quale mi dice che $x = x^+ - x^-$. Anzi, in realtà se posso scrivere x come $u - v$, con $u, v \geq 0$, se è $u \geq v$, ho che $x^+ = u - v$ e $x^- = v - v = 0$; similmente se $u \leq v$ si ha $x^+ = 0$ e $x^- = v - u$. Come si può notare, u, v non sono univocamente determinate da x).

Forma canonica:

$\max cx$
$Ax \leq b$
$x \geq 0$

Ogni problema di PL può essere messo in forma standard:

$\max cx$
$Ax = b$
$x \geq 0$

Significato di equivalenza dei problemi.

Variabili di slack e di surplus. E ricordato come si trattano le variabili "libere" (cioè senza vincolo di segno).

Dato un problema in forma canonica, ad esso possiamo associare con una procedura sistematica un problema in forma standard, che gli è equivalente. Nel senso che per ogni soluzione ottima del problema in forma canonica possiamo individuare una corrispondente soluzione ottima del problema in forma standard e viceversa.
Vediamo. Abbiamo un problema in forma canonica:
$\max cx$
$Ax \leq b$
$x \geq 0$
Introduciamo le variabili $y \in \mathbb{R}^m$ e consideriamo il seguente problema, in forma standard.
$\max cx + 0y$
$Ax + Iy = b$
$x \geq 0$
$y \geq 0$
Proviamo che, se $\overline{x}$ è una soluzione ottima del problema in forma canonica, allora $(\overline{x}, b - A\overline{x})$ è soluzione del problema in forma standard definito qui sopra.
E' immediato verificare che $(\overline{x}, b - A\overline{x})$ soddisfa i vincoli. Resta da provare che è soluzione ottima.
Supponiamo per assurdo che vi sia un elemento $(\hat{x}, \hat{y})$ che soddisfa anch'esso i vincoli: $A\hat{x} + I\hat{y} = b$
$\hat{x} \geq 0$
$\hat{y} \geq 0$
e t.c.:
$c\hat{x} + 0\hat{y} > c\overline{x} + 0(b - A\overline{x})$
Si vede subito che il vettore $\hat{x}$ soddisfa i vincoli del problema (in forma canonica) di partenza:
Da $A\hat{x} + I\hat{y} = b$, tenuto conto del fatto che $I\hat{y} = \hat{y}$ e $\hat{y} \geq 0$, abbiamo che $A\hat{x} \leq b$.
Inoltre, ovviamente è anche $\hat{x} \geq 0$.
Allora abbiamo un vettore $\hat{x}$ che soddisfa i vincoli del problema (in forma canonica) di partenza. Abbiamo anche supposto che $c\hat{x} + 0\hat{y} > c\overline{x} + 0(b - A\overline{x})$, ovvero che $c\hat{x} > c\overline{x}$. Ma allora $\hat{x}$ soddisfa i vincoli del problema in forma canonica e fa assumere alla funzione obiettivo un valore strettamente maggiore di quello che essa assume in $\hat{x}$. Contraddizione col fatto che $\hat{x}$ era una soluzione ottima del problema dato (in forma canonica).

E' anche possibile associare ad un problema in forma standard un equivalente problema in forma canonica. Qui la strada è più semplice, perché, dato:
$\max cx$
$Ax = b$
$x \geq 0$

sarà sufficiente considerare questo problema (ovviamente in forma canonica ed ovviamente equivalente al problema dato):
$\max cx$
$Ax \leq b$
$-Ax \leq -b$
$x \geq 0$

Interpretazione geometrica.
Supponiamo, per comodità, di avere un problema di PL in forma canonica.
Chiamiamo P l'insieme dei punti di $\mathbb{R}^n$ che soddisfano il vincolo.
Lo possiamo vedere come intersezione di un numero finito di semispazi.
Quindi è convesso (i semispazi sono convessi e la intersezione di convessi è convessa).
E' un convesso speciale. Non tutti i convessi sono intersezione di un numero finito di semispazi.
Visto che i convessi che sono intersezione di un numero finito di semispazi sono detti poliedri, P è un poliedro.
Un poliedro può essere limitato (in tal caso lo chiamiamo politopo) o illimitato. Se è limitato, può essere vuoto oppure no,
Ci interessano i punti estremi di P. Li chiameremo vertici. Di fatto, corrispondono all'idea elementare di vertice.
[Motzkin] Ogni poliedro P può essere decomposto così: $P = \mathrm{conv}(X) + \mathrm{cone}(Y)$, con X ed Y finiti (Y può essere vuoto; lo è se e solo se P è limitato).

Qui un esempio relativo al teorema di Motzkin: esempio per Motzkin, file pdf (una pagina).

Se P è non vuoto ed inoltre $P \subseteq \mathbb{R}^n_{\geq}$ e X è minimale, allora P ha vertici e i suoi vertici sono gli elementi di X.
Se P è non limitato, come detto Y è non vuoto. Allora, cx ha massimo su P se e solo se $cy \leq 0$ per ogni $y \in Y$.

ESERCIZI:
Provare, usando la decomposizione di Motzkin, che se la funzione obiettivo di un problema di PL è superiormente limitata su un vincolo non vuoto, allora ha massimo.

Ve 11 dicembre 2009, 3h-22h:
Noi lavoreremo con pb. in forma standard. Anzi, lavoreremo nelle ipotesi:
- $b_j \geq 0$ per ogni $j \in \{1, ..., m\}$.
- $m \leq n$
- $r(A) = m$

La prima condizione la possiamo soddisfare facilmente (con un cambiamento di segno)
Per le altre due condizioni, se la seconda non è soddisfatta lo vediamo facilmente.
La terza condizione non è di agevole verifica. Tuttavia, il metodo del simplesso ci permetterà, tra le altre cose, di verificare se è soddisfatta oppure no.

Interpretazione algebrica.
Supponiamo, per comodità, di avere un problema di PL in forma standard.
E anche che siano soddisfatte le condizioni: $b \geq 0$, $m \leq n$ e $r(A) = m$.
Prendiamo una sottomatrice $m \cdot m$ (cioè con m righe ed m colonne) B di A, che sia non-singolare.
La chiamiamo matrice di base (le sue colonne, e le variabili corrispondenti alle sue colonne le chiamiamo rispettivamente colonne e variabili di base).
Il sistema $Bx_B = b - F\overline{x}_F$ ha soluzione unica comunque fissiamo $\overline{x}_F$.
Prima osservazione.
Se fissiamo $\overline{x}_F = 0$, la soluzione che otteniamo è detta soluzione di base.
Non è detto che la soluzione di base $x_B = B^{-1}b$ soddisfi i vincoli di non negatività. Cioè non è detto che sia $B^{-1}b \geq 0$. Se questo è verificato, diremo che la soluzione è una soluzione *ammissibile* di base.
Si può dimostrare che, per un problema in forma standard, x è un vertice di P se e solo se è una soluzione ammissibile di base.
Nota: possono esserci più basi che danno luogo alla stessa soluzione ammissibile di base (e quindi allo stesso vertice).

Teorema fondamentale della PL Sia dato un problema di PL in forma standard. Indico con P il poliedro dei vincoli. Allora:
- se $P \neq \emptyset$, allora esiste una soluzione ammissibile di base (che, ricordo, corrisponde a un vertice di P)
- se esiste una soluzione ottima, allora esiste una soluzione ammissibile di base ottima.

Siamo allora sicuri che, se il problema di PL ha soluzione, tra le soluzioni ci sarà almeno una soluzione ottima di base, ovvero un vertice di P.
Nota: tutto questo vale se abbiamo un problema in forma standard (andrebbe bene anche se fosse in forma canonica). Se non è in forma standard, allora non è detto che P abbia vertici. Come esempi, vedansi gli esercizi già proposti.

Osservazione.
Data una base B, possiamo esprimere le variabili di base in funzione delle variabili fuori base. Andando a sostituire nella espressione della funzione obiettivo, troviamo una funzione lineare nelle sole variabili fuori base. (O, se si vuole, anche nelle variabili di base, ma con coefficienti nulli rispetto a queste). I coefficienti di questa nuova funzione lineare trovata li chiamiamo costi ridotti (la terminologia sarebbe appropriata ad un problema di minimo, a dire il vero...).
Si prova facilmente che una soluzione di base, che sia ammissibile, è soluzione ottima del problema se i coefficienti di costo ridotti sono tutti minori o uguali di zero.

Sia B una matrice $m \cdot m$ non singolare (matrice di base). Allora $x = B^{-1}b$ è soluzione di base. Se $x \geq 0$, abbiamo una soluzione ammissibile di base.
Possiamo scrivere: $x_B = B^{-1}b - B^{-1}Fx_F$
Da cui: $cx = c_B B^{-1} b + \left(c_F - c_B B^{-1} F\right) x_F = COST + \hat{c}x$
Dove $\hat{c} = \left[\hat{c}_B \ \hat{c}_F\right] = \left[0 \ \ c_F - c_B B^{-1} F\right]$ sono i cosiddetti "coefficienti di costo ridotti" (il termine "costi" ha senso se la funzione obiettivo rappresenta un costo. Ma in tal caso di solito si preferisce minimizzare la funzione obiettivo, anziché massimizzarla. Adotterò pertanto usualmente la dizione "coefficienti ridotti"). Notare che i coefficienti di costo ridotto corrispondenti alle variabili di base sono tutti *nulli*.
Teorema. Sia dato un problema di PL in forma standard, che soddisfa le 3 condizioni. Se per una soluzione ammissibile di base si ha $\hat{c} \leq 0$

(ovvero tutti i corrispondenti costi ridotti sono minori o uguali di zero), allora è una soluzione ottima.
Dimostrazione (banale). E' $cx = c_B B^{-1} b + \left(c_F - c_B B^{-1} F\right) x_F$. Abbiamo che, nella soluzione ammissibile di base, $x_F = 0$ e quindi $cx = c_B B^{-1} b$. Mentre per ogni punto del poliedro dei vincoli P ogni variabile assume un valore non-negativo (ovviamente per i vincoli di non-negatività che valgono per tutte le variabili). Ma allora, se i coefficienti ridotti sono minori o uguali di zero, il contributo dato dal termine $\left(c_F - c_B B^{-1} F\right) x_F$ è minore o uguale di zero. Il che quindi ci garantisce che il valore ottenuto nella soluzione ammissibile di base è maggiore o uguale di ogni altro valore che la funzione obiettivo assume nell'ortante non negativo e quindi, a fortiori, in P.

Un primo esempio (molto semplice!):
$\max 2x + 3y$
con i vincoli:
$x + y = 1$
$x, y \geq 0$

La matrice dei coefficienti è $(1, 1)$ ed ha due sottomatrici (1x1...) non singolari. Posso scegliere quindi sia la x che la y come variabili di base.
Se scelgo la x il vincolo diventa: $x = 1 - y$. Per $y = 0$ ottengo $x = 1$, quindi la mia soluzione di base è anche ammissibile.
Per quanto riguarda la funzione obiettivo e i coefficienti ridotti, ho esplicitato il vincolo rispetto alla variabile fuori base e posso andare a sostituire nella funzione obiettivo. che diventa:
$2(1 - y) + 3y$.
Ovvero: $2 + 1y$. Naturalmente $COST = 2$ e i coefficienti ridotti sono: 0 (ovviamente) per la x (è la variabile in base), e 1 per la y.
Se scegliamo y come variabile in base, troviamo: $y = 1 - x$, e la soluzione di base è $(x, y) = (0, 1)$ che quindi è anch'essa ammissibile.
Per la funzione obiettivo, abbiamo: $2x + 3(1 - x) = 3 - x$. I coefficienti ridotti sono $(-1, 0)$. Quindi abbiamo trovato la soluzione ottima.

Il metodo del simplesso attraverso un esempio.

$\max x_1 + x_2$
con i vincoli:
$$\begin{cases} 6x_1 + 4x_2 \leq 24 \\ 3x_1 - 2x_2 \leq 6 \\ x_1, \ \ x_2 \geq 0 \end{cases}$$

Convertiamo in forma standard aggiungendo variabili di slack x_3 e x_4:
$\max x_1 + x_2$
con i vincoli:
$$\begin{cases} 6x_1 + 4x_2 + x_3 \qquad\quad = 24 \\ 3x_1 - 2x_2 \qquad\quad + x_4 = 6 \\ x_1, \ \ x_2, \ \ x_3, \ \ x_4 \geq 0 \end{cases}$$

Abbiamo matrice A, con 6 sottomatrici $2 \cdot 2$ non singolari.
Otteniamo 6 soluzioni di base, di cui 4 ammissibili e 2 no.
Partiamo dalla base x_3, x_4.
Otteniamo soluzione ammissibile di base che è $(0, 0, 24, 6)$, corrispondente al punto "A" nel disegno fatto alla lavagna, ovvero all'origine degli assi nel problema originario, in due variabili.
I costi ridotti sono uguali a quelli "normali" (ciò è dovuto al fatto che la matrice di base è l'identità e al fatto che sto usando come variabili di base tutte e sole le variabili di slack): $\hat{c} = (1, 1, 0, 0)$.
Se fosse $\hat{c} \leq 0$, STOP: abbiamo una soluzione ottima. Nel nostro esempio, non è $\hat{c} \leq 0$. Allora proviamo a cambiare base. Mettiamo in base una delle variabili corrispondente a un coefficiente di costo ridotto strettamente positivo.
Possiamo mettere in base x_1 oppure x_2. Scegliamo di mettere in base x_1.
Naturalmente, dovremo togliere una delle variabili che erano in base. Riscriviamo il sistema dei vincoli (mi riferisco ai vincoli $Ax = b$) tenendo uguale a zero le variabili che erano fuori base e che non facciamo entrare in base a questo passo (nel nostro esempio si tratta di x_2):
$+x_3 \qquad\quad = 24 - 6x_1$
$\qquad\quad +x_4 = 6 - 3x_1$
Se prendiamo $x_1 = \dfrac{24}{6}$ si annulla x_3. Ma x_4 sarà negativa e quindi non avremo una soluzione ammissibile di base. (Nel disegno fatto alla lavagna troviamo il punto "H")
Se prendiamo $x_1 = \dfrac{6}{3}$ si annulla x_4. E la x_3 resta non negativa. Quindi rimaniamo con una soluzione *ammissibile* di base.
Disegni e considerazioni relativi all'esempio: esempio PL, simplesso, file pdf (tre pagine).

In pratica, in generale, dobbiamo guardare i coefficienti *positivi* che compaiono nella colonna della matrice sotto la variabile che vogliamo far entrare in base. Tra le righe corrispondenti a questi coefficienti cerchiamo la riga per cui il quoziente fra il "termine noto" e il relativo coefficiente nella matrice è minimo. La variabile da far uscire dalla base è quella corrispondente a questa riga (se vi fossero più righe che rendono minimo il rapporto, scegliamo una tra queste, a nostro piacimento).
Se per caso tutti i coefficienti della matrice nella colonna della variabile che abbiamo deciso di far entrare in base fossero minori o uguali a zero? Vorrebbe dire che la funzione obiettivo non è superiormente limitata. Quindi, STOP: il problema dato non ha soluzioni ottime.
A questo punto si effettua la operazione di "pivoting" per far sì che la matrice corrispondente alle nuove variabili di base sia la matrice identità.
Da qui si riparte per il passo successivo. E si va avanti fino a che non si arriva allo STOP (che può esserci perché abbiamo trovato una soluzione ottima oppure perché sappiamo che la funzione obiettivo non è superiormente limitata).

Resta un dubbio da sciogliere: siamo sicuri che questo algoritmo termini dopo un numero finito di passi? La risposta è no. Saremmo sicuri che l'algoritmo termina dopo un numero finito di passi se sapessimo che non si ripassa mai da un vertice già "visitato". Questa eventualità non la possiamo escludere. Può verificarsi un caso di "cycling", che può occorrere in corrispondenza di soluzioni di base degeneri (ovvero con qualche coordinata nulla; questo avviene ad esempio se si trova che (almeno) due variabili rendono minimo il rapporto che ci serviva per trovare la variabile da far uscire dalla base). Il problema del "cycling" è più di carattere teorico che pratico. Comunque, vi è una regola (di Bland) che può essere usata se si vuole eliminare questo rischio.

DOCUMENTI:
Degenerazione:
pagine da Dantzig (fig. uno e due e breve testo)
pagine da Tadei e Della Croce (fig. uno due e tre).

ESERCIZI:
Esercizio 1.
Si consideri il problema $\max x + y$, con i vincoli $x \geq 2$ e $x \leq 3$. Trasformarlo in forma canonica e, in questa forma, trovare i vertici del poliedro dei vincoli. Idem con la forma standard.
Soluzione.
$\max x + y$
$x \geq 2$
$x \leq 3$
Trasformato in forma canonica:
$\max u - v + s - t$
$-u + v \leq -2$
$u - v \leq 3$
$u, v, s, t \geq 0$
Il poliedro in $\mathbb{R}^4$ ha come vertici $(3, 0, 0, 0), (2, 0, 0, 0)$.
Lo trasformiamo in forma standard:
$\max u - v + s - t$
$-u + v + h = -2$
$u - v + k = 3$
$u, v, s, t, h, k \geq 0$
La matrice dei coefficienti è:

$$\begin{array}{c} \begin{array}{cccccc} u & v & s & t & h & k \end{array} \\ \begin{pmatrix} -1 & 1 & 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

Vi sono 15 sottomatrici, di cui 5 non singolari: uh, uk, vh, vk, hk (il senso dei simboli dovrebbe essere ovvio). Abbiamo in loro corrispondenza 5 soluzioni di base (di cui due ammissibili):
$(3, 0, 0, 0, 1, 0)$, $(2, 0, 0, 0, 0, 1)$, $(0, -3, 0, 0, 1, 0)$, $(0, -2, 0, 0, 0, 1)$, $(0, 0, 0, 0, -2, 3)$.

Esercizio 1.
Si consideri il problema $\max x + y$, con i vincoli $x + y \geq 2$ e $x + y \leq 3$. Trasformarlo in forma canonica e, in questa forma, trovare i vertici del poliedro dei vincoli. Idem con la forma standard. Osservare che, sia in forma canonica che in forma standard, assume un massimo in un vertice. A cosa corrisponde questo vertice, nel problema originario?

Abbiamo visto che: $cx = c_B B^{-1} b + \left(c_F - c_B B^{-1} F\right) x_F = COST + \hat{c}x$. Domanda: come mai la funzione obiettivo, che è lineare, diventa affine? Ovvero, come mai spunta fuori la "COST"?

Ma 15 dicembre 2009, 2h-24h:
Riprendiamo l'esempio già analizzato, per vedere come si possa in pratica mettere in opera il metodo del simplesso.
Abbiamo:

$\max x_1 + x_2$
con i vincoli:

$$\begin{cases} 6x_1 + 4x_2 + x_3 = 24 \\ 3x_1 - 2x_2 + x_4 = 6 \\ x_1, \ x_2, \ x_3, \ x_4 \geq 0 \end{cases}$$

I conti fatti usando algebra lineare e geometria analitica possono essere "riassunti" lavorando su una tabella (nel gergo della PL, "tableau").
Vediamo il tableau:

```
        x_1   x_2   x_3   x_4

         1     1     0     0      0

x_3      6     4     1     0     24
x_4      3    -2     0     1      6
```

Nella prima riga indico i nomi delle variabili (in realtà questa riga serve a ben poco, può essere omessa).
Nella seconda riga indico i coefficienti ridotti corrispondenti alle variabili in base scelte. Nel nostro caso, essi coincidono con i coefficienti originari di costo (avendo messo "in base" le variabili di slack); l'ultimo elemento è il valore della funzione obiettivo, calcolato nella soluzione ammissibile che si sta considerando e *cambiato di segno*. Visto che nel nostro caso è zero, il fatto che ne dobbiamo cambiare il segno si nota poco...
Nella terza riga indico: il nome della prima variabile in base: (x_3), i coefficienti della prima riga della matrice A, il termine noto (b_1)
Nella quarta riga indico: il nome della seconda variabile in base: (x_4), i coefficienti della seconda riga della matrice A, il termine noto (b_2)

Ho due coefficienti ridotti positivi. Scelgo di far entrare in base x_1. Calcolo $\frac{24}{6} = 4$ e $\frac{6}{3} = 2$ per scoprire chi fare uscire. Esce x_4 perché è sulla sua riga che trovo il quoziente più piccolo.
Quindi entra x_1 ed esce x_4.
Operazione di pivoting per fare comparire una colonna (sotto ad x_1) di tutti zeri (compresi i coefficienti della funzione obiettivo) tranne che in

corrispondenza del pivot dove dovrà comparire 1 .
Divido per 3 la riga corrispondente a x_4 (notare che le operazioni le faccio anche sui "termini noti") per far comparire 1 in corrispondenza del pivot:

```
      x_1   x_2   x_3   x_4

       1     1     0     0      0

x_3    6     4     1     0     24
x_1    1   -2/3    0    1/3    6/3
```

Moltiplico per -2 la riga del pivot e la sommo alla riga x_3 in modo che compaia uno zero:

```
      x_1   x_2   x_3   x_4

       1     1     0     0      0

x_3    0     8     1    -2     12
x_1    1   -2/3    0    1/3     2
```

Moltiplico per $-\frac{1}{3}$ la riga del pivot e la sommo alla riga dei coefficienti c in modo che compaia uno zero:

```
      x_1   x_2   x_3   x_4

       0    5/3    0    -1/3   -2

x_3    0     8     1    -2     12
x_1    1   -2/3    0    1/3     2
```

Ho ottenuto un nuovo tableau in forma canonica, questa volta per la base x_1, x_3:

```
      x_1   x_2   x_3   x_4

       0    5/3    0    -1/3   -2

x_3    0     8     1    -2     12
x_1    1   -2/3    0    1/3     2
```

Notare il modo in cui ho contrassegnato le righe!
Ho ancora un coefficiente ridotto positivo, quindi faccio entrare in base la variabile corrispondente x_2). Visto che nella colonna di x_2 c'è solo un elemento strettamente positivo, sarà la variabile di base corrispondente a tale riga (e quindi x_3) ad uscire di base.
Avendo individuato il pivot, ricalcolo il nuovo tableau. Etc.

Metodo a due fasi per determinare una soluzione ammissibile di base.
Il problema ausiliario di PL ha come funzione obiettivo $\sum_{i=1}^{m} -y_i$. Per questo problema l'insieme dei vincoli, il poliedro P, è sempre non vuoto. Inoltre, la funzione obiettivo è superiormente limitata da 0 e quindi esiste sempre soluzione ottima. Se il valore della funzione obiettivo è 0, allora il problema di partenza aveva una soluzione ammissibile di base (e il metodo del simplesso, applicato al problema ausiliario, ce la fa trovare). Se invece il valore della funzione obiettivo è strettamente minore di zero, allora il problema dato non ha soluzione ammissibile di base.

Esempio.
In sintesi: un esempio di applicazione del metodo a due fasi. Determinazione dei coefficienti ridotti per il problema ausiliario. La soluzione del problema ausiliario non ha in base le variabili ausiliarie e quindi ci dà direttamente una soluzione ammissibile di base del problema originario. Calcolo dei coefficienti di costo ridotti per il problema originario.

L'esempio in dettaglio.
Consideriamo il problema:
$\max 4x_1 + x_2 + x_3$
$2x_1 + x_2 + 2x_3 = 4$
$3x_1 + 3x_2 + x_3 = 3$
$x_1, x_2, x_3 \geq 0$
Non è in forma canonica, ma è direttamente in forma standard. Questo non ci permette di utilizzare come base ammissibile di partenza quella costituita dalla variabili di slack. C'è il problema di trovare un "punto di partenza". Ovviamente, data la semplicità del problema, questo si può fare "a mano" e per questo rinvio al disegno.
Utilizziamo il metodo a due fasi. Cerchiamo:
$\max -y_1 - y_2$
$2x_1 + x_2 + 2x_3 + y_1 = 4$
$3x_1 + 3x_2 + x_3 + y_2 = 3$
$x_1, x_2, x_3, y_1, y_2 \geq 0$
Proviamo a scrivere il tableau iniziale:

```
      x_1   x_2   x_3   y_1   y_2

       0     0     0    -1    -1       <<<  attenzione!

y_1    2     1     2     1     0     4
y_2    3     3     1     0     1     3
```

Ho scritto "attenzione!" perché i numeri riportati su quella linea NON sono i coefficienti ridotti. Dobbiamo esprimere le variabili in base in funzione di quelle fuori base e poi andare a sostituire i coefficienti così trovati:
$-y_1 = 2x_1 + x_2 + 2x_3 + y_1 - 4$
$-y_2 = 3x_1 + 3x_2 + x_3 + y_2 - 3$
Sommando membro a membro otteniamo:
$-y_1 - y_2 = 5x_1 + 4x_2 + 3x_3 - 7$
Abbiamo trovato (per le vie brevi) i coefficienti ridotti che mettiamo nella prima riga:

```
        x_1   x_2   x_3   y_1   y_2

  7      5     4     3     0     0

 y_1     2     1     2     1     0     4
 y_2     3     3     1     0     1     3
```

Notare "7" che indica il valore (cambiato di segno) della funzione obiettivo nella base ammissibile trovata.
Visto che i coefficienti di costo ridotto sono positivi (tutti!) cambiamo base. Mettiamo in base x_1. Calcoliamo $\Theta = \min\left\{\frac{4}{2}, \frac{3}{3}\right\} = 1$, e il min è ottenuto in corrispondenza di y_1 che quindi facciamo uscire di base.
Per il tableau, cominciamo a dividere la riga pivot per 3, in modo da ottenere 1 :

```
        x_1   x_2   x_3   y_1   y_2

  7      5     4     3     0     0

 y_1     2     1     2     1     0     4
 x_1     1     1    1/3    0    1/3    1    <<<  riga modificata
```

Sottraggo alla prima riga la riga pivot (già modificata) moltiplicata per 2 (il che è equivalente a sottrarre la riga pivot originaria moltiplicata per $\frac{2}{3}$:

```
        x_1   x_2   x_3   y_1   y_2

  7      5     4     3     0     0

 y_1     0    -1    4/3    1   -2/3    2    <<<  riga modificata
 x_1     1     1    1/3    0    1/3    1    <<<  riga modificata
```

C'è ancora da modificare la riga della funzione obiettivo. Dobbiamo fare apparire 0 al posto del 5 che abbiamo in corrispondenza della colonna pivot. Sottraggo la riga pivot (già modificata) moltiplicata per 5 (ovvero, quella originaria moltiplicata per $\frac{5}{3}$).
Otteniamo:

```
        x_1   x_2   x_3   y_1   y_2

  2      0    -1    4/3    0   -5/3

 y_1     0    -1    4/3    1   -2/3    2
 x_1     1     1    1/3    0    1/3    1
```

Come si vede, abbiamo ancora un coefficiente ridotto positivo. Facciamo entrare in base x_3. E' $\Theta = \min\left\{\frac{2}{\frac{4}{3}}, \frac{1}{\frac{1}{3}}\right\} = \min\left\{\frac{3}{2}, 3\right\} = \frac{3}{2}$. Quindi esce di base y_1 (cosa che ci fa piacere: a questo punto tutte e due le variabili "ausiliarie" sono uscite di base!). Il nuovo tableau (omettiamo i conti intermedi) è:

```
        x_1   x_2   x_3   y_1   y_2

  0      0     0     0    -1    -1

 x_3     0   -3/4    1    3/4 -1/2   3/2
 x_1     1    5/4    0   -1/4  1/2   1/2
```

Visto che tutti i coefficienti ridotti sono minori o uguali a zero, abbiamo trovato una soluzione ottima per il problema artificiale, che è $x_1 = \frac{1}{2}$ e $x_3 = \frac{3}{2}$, come si vede subito "a occhio" dal tableau finale; ovviamente i valori delle tre variabili fuori base è zero.
Non solo: il valore della funzione obiettivo è 0 e quindi esiste una soluzione ammissibile di base per il problema originario.
Come già notato, non è rimasta nessuna variabile ausiliaria in base, quindi abbiamo già automaticamente una soluzione ammissibile di base per il problema dato.
Ecco il tableau:

```
        x_1   x_2   x_3

  0      0     0     0         <<< attenzione!

 x_3     0   -3/4    1    3/2
 x_1     1    5/4    0    1/2
```

Nella riga dei coefficienti dobbiamo mettere i coefficienti ridotti, che otteniamo dal sistema originario:
$2x_1 + x_2 + 2x_3 = 4$
$3x_1 + 3x_2 + x_3 = 3$
Questo sistema (lo vediamo dal tableau che abbiamo sotto mano) è equivalente a: $0 - \frac{3}{4}x_2 + x_3 = \frac{3}{2}$
$x_1 + \frac{5}{4}x_2 + 0 = \frac{1}{2}$
Da qui è immediato esprimere x_1 e x_3 in funzione di x_2 per trovare i coefficienti ridotti:
$x_3 = \frac{3}{2} + \frac{3}{4}x_2$
$x_1 = \frac{1}{2} - \frac{5}{4}x_2$
Allora $c \dot{x} = 4x_1 + x_2 + x_3 = 4\left(\frac{1}{2} - \frac{5}{4}x_2\right) + x_2 + \left(\frac{3}{2} + \frac{3}{4}x_2\right) = \frac{7}{12} - \frac{13}{4}x_2$.
Allora il tableau iniziale è:

```
     x_1  x_2  x_3

      0 -13/4   0   -7/2

x_3   0  -3/4   1    3/2
x_1   1   5/4   0    1/2
```

Abbiamo avuto fortuna.Il coefficiente ridotto è minore o uguale a zero. Quindi abbiamo già trovato la soluzione ottima, che è. $\left(\frac{1}{2}, 0, \frac{3}{2}\right)$. E il valore ottimo che è $\frac{7}{2}$. Cose che già sapevamo dai conti a mano...

DOCUMENTI:
Degenerazione:
pagine da Dantzig (fig. uno e due e breve testo)
pagine da Tadei e Della Croce (fig. uno due e tre).
.
ESERCIZI:
.

Gi 25 febbraio 2010, 3h-27h:
Un esempio con funzione obiettivo non superiormente limitata.
$\max x_1 + x_2$
$x_2 \leq x_1 + 1$
$x_1, x_2 \geq 0$
Il poliedro P dei vincoli è la porzione di primo quadrante che sta sotto la retta $x_2 = x_1 + 1$.
Il teorema di Motzkin mi dice che posso scrivere: $P = \mathrm{conv}(X) + \mathrm{cone}(Y)$, con X ed Y finiti. Come X posso prendere $\{(0,0), (0,1)\}$. Come Y, ad esempio $\{(1,0), (1,1)\}$. Sappiamo che il nostro problema sarà superiormente limitato (e quindi avrà max, grazie alle proprietà caratteristiche dei problemi di programmazione lineare) se e solo se $c \cdot y \leq 0$ per ogni $y \in Y$ (c è il vettore dei coefficienti della funzione obiettivo).
La dim di questo fatto può essere fatta così: ogni $x \in P$ è esprimibile come $x = x^c + x^k$ ($x^c \in \mathrm{conv}(X)$, $x^k \in \mathrm{cone}(Y)$). Ma $\exists\, C \in \mathbb{R}$ t.c. $|x^c| \leq C$ per ogni $x^c \in \mathrm{conv}(X)$ (è un insieme limitato). Allora: $c \cdot x = c x^c + x^k = c \cdot x^c + c \cdot x^k \leq ||c||\, C$ se $c \cdot y \leq 0$ per ogni $y \in Y$
(Per comodità, visto che Y è finito, posso assumere che sia $Y = \{y_1, ..., y_m\}$. Si ha che $y \in \mathrm{cone}(Y)$ se e solo se $x^k = \sum_{j=1}^{m} \mu_j y_j$, con $\mu_j \geq 0$; pertanto $c \cdot x^k = c \cdot \sum_{j=1}^{m} \mu_j y_j = \sum_{j=1}^{m} c \cdot \mu_j y_j$, con tutti gli addendi minori o uguali di zero). E, quindi, la funzione obiettivo è superiormente limitata.
Viceversa, se non si ha $c \cdot y \leq 0$ per ogni $y \in Y$, vuol dire che $c \cdot \bar{y} > 0$ per un $\bar{y} \in Y$. Ma allora, prendiamo gli elementi di P del tipo $\bar{x} + \mu \bar{y}$ (con $\bar{x}$ scelto a piacere in X): $c \cdot \bar{x} + \mu \bar{y} = c \dot{\bar{x}} + c \cdot \mu \bar{y}$ che va a $-\infty$ per μ tendente a $+\infty$.
Nel nostro esempio, avendo $c = (1,1)$, ed essendo $(1,1) \cdot (1,0) > 0$, siamo certi che il problema dato non ha soluzioni in quanto la funzione obiettivo non è superiormente limitata (cosa che si vedeva ad "occhio", d'altronde...).
Vediamo se, usando il metodo del simplesso, ci accorgiamo del fatto che la funzione obiettivo non è superiormente limitata. Dapprima trasformiamo il problema in forma standard:
$\max x_1 + x_2$
$-x_1 + x_2 + x_3 = 1$
$x_1, x_2, x_3 \geq 0$
Il tableau iniziale è:

```
     x_1  x_2  x_3

      1    1    0    0

x_3  -1    1    1    1
```

Si noti che abbiamo un coefficiente ridotto (quello di x_1) che è positivo, e sotto abbiamo che tutti i coefficienti (a dire il vero questi "tutti" sono uno solo, in questo problema banale...) sono minori o uguali di zero. Ciò mi garantisce che, se volessi far entrare in base x_1, non avrei alcuna limitazione e quindi (essendo il coefficiente ridotto positivo) potrei far aumentare a mio piacimento il valore della funzione obiettivo. D'altronde, è ciò che si vede bene dal disegno: se dalla soluzione ammissibile di base (l'origine) mi sposto nella direzione dell'asse x_1, non incontro alcuna limitazione. E, sull'asse delle x_1, i valori della funzione obiettivo crescono senza limitazione superiore.
Un esempio di soluzione di base degenere. Un passo dell'algoritmo del simplesso non fa spostare dal vertice in cui si era.

Dualità: duale di un problema di PL in forma canonica.
Dato:
$\max cx$
t.c. $\begin{cases} Ax \leq b \\ x \geq 0 \end{cases}$
Il duale è:
$\min yb$
t.c. $\begin{cases} yA \geq c \\ y \geq 0 \end{cases}$
Il duale del duale è il problema di partenza (detto primale).

Perché è naturale vedere c come "vettore riga", ovvero come matrice $1 \cdot n$ (è la matrice di rappresentazione di una applicazione lineare da $\mathbb{R}^n$ in $\mathbb{R}$).

Duale di un problema in forma standard.
Teorema debole della dualità ($cx \leq yb$ per x, y soluzioni ammissibili, rispettivamente per problema primale e duale) e dimostrazione.
Conseguenze di questo teorema.

Teorema forte della dualità. Se (P) ha soluzione ottima, allora:
- anche (D) ha soluzione ottima e max(P) = min(D)
- se $x^\star$ è soluzione ottima per (P), allora esiste $y^\star$ ottima per (D) e $cx^\star = y^\star b$.

Tabella riassuntiva di alcune conseguenze dei due teoremi di dualità

P \ D	vuoto	sup. illimit.	sol. ott.
vuoto	SI	SI	NO
sup. illimit.	SI	NO	NO
sol. ott.	NO	NO	SI

Un esempio di problemi duali. Disegno delle zone ammissibili.
Il primale è quello già considerato più volte:

$\max x_1 + x_2$
con i vincoli:
$$\begin{cases} 6x_1 + 4x_2 \leq 24 \\ 3x_1 - 2x_2 \leq 6 \\ x_1, \ \ x_2 \geq 0 \end{cases}$$

Il suo duale è:

$\min 24y_1 + 6y_2$
con i vincoli:
$$\begin{cases} 6y_1 + 3y_2 \geq 1 \\ 4y_1 - 2y_2 \geq 1 \\ y_1, \ \ y_2 \geq 0 \end{cases}$$

Già sappiamo che la soluzione ottima del problema primale è $x_1 = 0, x_2 = 6$. E' immediato verificare che la soluzione ottima del duale è $y_1 = \frac{1}{4}, y_2 = 0$, e che il valore sia del primale che del duale è 6.
Qui un disegno del primale e del duale.

DOCUMENTI:
.
ESERCIZI:
.

Lu 1 marzo 2010, 2h-29h:
Il metodo del simplesso applicato al problema duale di:

$\max x_1 + x_2$
con i vincoli:
$$\begin{cases} 6x_1 + 4x_2 \leq 24 \\ 3x_1 - 2x_2 \leq 6 \\ x_1, \ \ x_2 \geq 0 \end{cases}$$

Ovvero, a:

$\min 24y_1 + 6y_2$
con i vincoli:

$$\begin{cases} 6y_1 + 3y_2 \geq 1 \\ 4y_1 - 2y_2 \geq 1 \\ y_1,\ \ y_2 \geq 0 \end{cases}$$

Consideriamo il duale e trasformiamolo in forma standard. Abbiamo:

$\max -24y_1 + -6y_2$
con i vincoli:

$$\begin{cases} 6y_1 + 3y_2 - y_3 \qquad = 1 \\ 4y_1 - 2y_2 \qquad -y_4 = 1 \\ y_1,\ \ y_2,\ \ y_3,\ \ y_4 \geq 0 \end{cases}$$

Volendo applicare il metodo del simplesso, non possiamo prendere come variabili in base le variabili di slack, visto che l'origine non sta nell'insieme descritto dai vincoli.
Ma dalla figura sappiamo che $(y_1, y_2) = \left(\frac{1}{4}, 0\right)$ ci danno le prime due coordinate di una soluzione ammissibile di base (infatti il punto $\left(\frac{1}{4}, 0\right)$ è un vertice del poliedro dei vincoli; anzi, è l'unico vertice). Per trovare le altre due coordinate, sostituisco nelle equazioni dei vincoli:

$$\begin{cases} 6\frac{1}{4} + 3 \cdot 0 - y_3 \qquad = 1 \\ 4\frac{1}{4} - 2 \cdot 0 \qquad -y_4 = 1 \end{cases}$$

Da cui ricaviamo immediatamente che $y_2 = \frac{1}{2}$ e $y_4 = 0$.
Allora, abbiamo in base le variabili y_1 e y_3, mentre y_2 e y_4 sono fuori base.
La matrice dei coefficienti è:

```
6   3  -1   0
4  -2   0  -1
```

La matrice di base B è:

```
6  -1
4   0
```

Come si vede, B non è la matrice identità. Questo significa che non abbiamo un tableau in forma canonica. Lo trasformiamo in forma canonica moltiplicando "tutto" per B^{-1}.
Con conti standard otteniamo che B^{-1} è:

```
 0  1/4
-1  3/2
```

Otteniamo allora che $B^{-1}F$ è:

```
-1/2  -1/4
 -6   -3/2
```

I "termini noti" saranno $B^{-1}b$:

```
1/4
1/2
```

Calcoliamo i coefficienti ridotti. La funzione obiettivo si può esprimere così: $z = c_B B^{-1} b + \left(c_F - c_B B^{-1} F\right) x_F$, che nel nostro caso ci dà $z = -6 - 6y_2 - 6y_4$. Quindi abbiamo sia il valore della funzione obiettivo nella soluzione ammissibile di base di cui ci stiamo occupando: esso è 6.
Possiamo allora scrivere il tableau (in forma canonica):

```
     y_1  y_2  y_3  y_4

      0   -6    0   -6     6

y_1   1  -1/2   0  -1/4   1/4
y_3   0   -6    1  -3/2   1/2
```

Come doveva essere, i coefficienti ridotti sono tutti minori o uguali a zero. Quindi la nostra soluzione ammissibile di base è anche ottima (come già ben sapevamo).
Duale di un modello di trasporto: interpretazione.
Ricordo che un problema di trasporto (caso "bilanciato", ovvero quando $\sum_{i=1}^{n} a_i = \sum_{j=1}^{m} b_j$) è:

$\min \sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij} x_{ij}$
con i vincoli:

$\sum_{j=1}^{m} x_{ij} = a_i$ per $i = 1, \ldots, n$
$\sum_{i=1}^{n} x_{ij} = b_j$ per $j = 1, \ldots, m$
$x_{ij} \geq 0$ per $i = 1, \ldots, n$, $j = 1, \ldots, m$

I dati possono essere interpretati nel modo seguente (riferiti ad una azienda):
a_i quantità prodotta dalla fabbrica i
b_j capacità del magazzino j
c_{ij} costo di trasporto unitario da i a j

Il duale è (si noti che, avendo nel primale vincoli di uguaglianza, le variabili duali sono "libere", ovvero non hanno vincoli di segno):

$\max - \sum_{i=1}^{n} a_i \lambda_i + \sum_{j=1}^{m} b_j \mu_j$
con i vincoli
$-\lambda_i + \mu_j \leq c_{ij}$ per $i = 1, \ldots, n$, $j = 1, \ldots, m$

Possibile interpretazione (esternalizzazione dei servizi interni?): un intermediario si offre di:
- comprare da i al prezzo (unitario) λ_i
- vendere a j al prezzo (unitario) μ_j

Il vincolo esprime il fatto che deve essere competitivo rispetto ai costi dell'azienda.
E la funzione obiettivo indica che l'intermediario vuole massimizzare il suo guadagno.

Problema della dieta e suo duale.

Condizioni di ortogonalità (o "complementarity slackness").

Le seguenti tre condizioni sono equivalenti.

(1)
$x^\star$ e $y^\star$ soluzioni di (P) e (D) (problema primale e duale, in forma canonica).

(2)
$Ax^\star \leq b$ e $x^\star \geq 0$ (ammissibilità per il problema primale)
$y^\star A \geq c$ e $y^\star \geq 0$ (ammissibilità per il problema duale)
$cx^\star = y^\star b$

(3)
$y^\star \cdot (b - Ax^\star) = 0$ (condizione di ortogonalità, o di "complementarity slackness")
$(y^\star A - c) \cdot x^\star = 0$ (condizione di ortogonalità, o di "complementarity slackness")
$Ax^\star \leq b$ e $x^\star \geq 0$ (ammissibilità per il problema primale)
$y^\star A \geq c$ e $y^\star \geq 0$ (ammissibilità per il problema duale)

Per provare l'equivalenza fra (1), (2) e (3), proverò che (2) $\Rightarrow$ (1), (1) $\Rightarrow$ (3) e che (3) $\Rightarrow$ (2).
Osservo comunque che è immediato provare che (1) implica (2): ce lo dice il teorema di dualità forte.

(2) $\Rightarrow$ (1)
Ricordiamo che il teorema di dualità debole ci dice che:
$cx \leq y^\star b$ per ogni soluzione ammissibile x. Quindi, $y^\star b$ è un maggiorante per la funzione obiettivo (del primale). Visto che $x^\star$ è tale che $cx^\star = y^\star b$, ovvero la funzione obiettivo assume un valore uguale a quello di un maggiorante, ne segue che $x^\star$ è soluzione ottima per (P).
Analogamente si dimostra che $y^\star$ è soluzione ottima per (D).

(1) $\Rightarrow$ (3)
Ricordiamo come nella dimostrazione del teorema di dualità debole si provi che $cx \leq yAx \leq yb$, per ogni soluzione ammissibile x di (P) ed y di (D). Quindi vale anche, in particolare, che: $cx^\star \leq y^\star Ax^\star \leq y^\star b$
Ma $cx^\star = y^\star b$ per il teorema di dualità forte. Quindi $cx^\star = y^\star Ax^\star = y^\star b$, da cui le condizioni di ortogonalità.

(3) $\Rightarrow$ (2)
Immediato.

Osservazione importante sulle condizioni di ortogonalità. E':
$y^\star(b - Ax^\star) = 0$ e $(y^\star A - c)x^\star = 0$
Ma y_j e $(b - Ax^\star)_j$ sono maggiori o uguali di zero. Quindi, ognuno degli addendi deve essere nullo.
Quindi, se ad esempio fosse $(y^\star)_{j_0} > 0$, avremmo $(b - Ax^\star)_{j_0} = 0$.

Verifica di queste condizioni in un esempio (il solito).

Tre condizioni equivalenti di ottimalità per primale e duale: la definizione, l'uguaglianza dei valori, le condizioni di complementarità.
Riguardo alle condizioni di complementarità, un'avvertenza. Usiamo il solito esempio per capire dove si può sbagliare.
Le soluzioni sono: $(0, 6, 0, 18)$ per il primale e $\left(\frac{1}{4}, 0, \frac{1}{2}, 0\right)$ per il duale.
Si può notare che a coordinate positive della soluzione del primale (seconda e quarta coordinata) corrispondono coordinate nulle della soluzione del duale (seconda e quarta coordinata). Analogamente a coordinate positive nella soluzione ottima del duale (prima e terza

coordinata) corrispondono coordinate nulle (prima e terza) nella soluzione ottima del primale.
Quanto detto è vero, ma **non c'entra nulla con la condizione di ortogonalità**. E' un **fatto accidentale** di questo problema ed è **gravemente fuorviante**.
Ricordiamo che il problema originario era dato in forma canonica e l'abbiamo trasformato in forma standard mediante l'uso delle variabili di slack. Quindi, partiti da un problema n-dimensionale, abbiamo ottenuto un problema (n+m)-dimensionale. Per il duale, abbiamo fatto una cosa simile: partendo da un problema in forma canonica (m-dimensionale), lo abbiamo trasformato in forma standard ottenendo un problema di dimensione m+n.
La corretta corrispondenza tra variabili è tra i vettori (1,...x...,n,1,...slack...m) e (1,...y...,m,1,...slack...n), intesa nel senso seguente:
(1,...x...n) $<\!\!\rightarrow$ (1...slack...n)
e
(1,...y...m) $<\!\!\rightarrow$ (1...slack...m)
Pertanto, nel nostro esempio, le condizioni di complementarità ci garantiscono che da $x_2 > 0$ segue che $y_4 = 0$, non che $y_2 = 0$ (anche se è vero...). E così via...

DOCUMENTI:
Per il duale di un modello di trasporto, vedasi: appunti PL da Pisa [originali da qui: http://www.di.unipi.it/optimize/], a pag. 18 del pdf. Contiene anche l'interpretazione del duale del problema della dieta (a pag. 17 del pdf).
ESERCIZI:
.
OSSERVAZIONE
Nell'esempio del problema di trasporti, la condizione che il sistema fosse "bilanciato", ovvero che sia $\sum_{i=1}^{n} a_i = \sum_{j=1}^{m} b_j$, non è né ovvia né innocua. Questa condizione presuppone di essere in una condizione di equilibrio, ovvero che la capacità produttiva totale sia uguale alla capacità totale di stoccaggio.

Gi 4 marzo 2010, 3h-32h:
Tabellina di corrispondenze fra primale e duale (da Tadei e Della Croce, vedasi anche Hiller e Lieberman):

Analisi di sensitività: breve introduzione di carattere generale (invito a comprendere per bene, calcoli inclusi, il caso più banale: come dipendono dai parametri a e b le soluzioni dell'equazione $ax + b = 0$) e poi studio specifico della dipendenza dal "secondo membro" dei vincoli, b.

Si può fare una analisi di *sensitività* della soluzione della equazione $\nabla f(x) = 0$ rispetto a parametri. Se abbiamo funzioni di classe C^2 possiamo scrivere una condizione (vedi Bertsekas). Nel caso semplice, di una funzione di una variabile, dipendente da un parametro a, ovvero data $f(x, a)$, abbiamo: $\sigma'(a) = -\frac{\frac{\partial^2 f}{\partial a \partial x}(\sigma(a), a)}{\frac{\partial^2 f}{\partial x^2}(\sigma(a), a)}$.
Ho indicato con $\sigma'(a)$ la derivata della funzione $a \mapsto \sigma(a)$ che ad ogni valore del parametro a associa il valore, supposto unico, del punto critico. La si trova (se sono soddisfatte le opportune condizioni) da $\frac{d}{dx} f(x, a) = 0$, usando il teorema delle funzioni implicite (di Dini).
Si può fare una "verifica" con $f(x, a) = x^2 - 6ax$. E' $\frac{d}{dx} f(x, a) = 2x - 6a$. Il punto che annulla il gradiente (qui abbiamo un punto di minimo) è $x = 3a$. Ovvero, $\sigma(a) = 3a$ E quindi $\sigma'(a) = 3$. Il che significa che, se il parametro a varia da $\bar{a}$ a $\bar{a} + \Delta a$, il punto critico si sposta, passando da $\bar{x}$ a $\bar{x} + 3\Delta a$ (ovviamente $\bar{x}$ è il punto critico che si ottiene in corrispondenza del valore $\bar{a}$ del parametro).

Sensitività e interpretazione dei moltiplicatori di Lagrange. Per i dettagli delle ipotesi ed una formulazione corretta del risultato, vedi Bertsekas. Qui ci si limita a mettere l'accento sulla "formula". Se consideriamo il problema di minimizzare $f(x)$ con il vincolo (vettoriale: $u \in \mathbb{R}^m$) $h(x) = u$, abbiamo che $\nabla p(u) = -\lambda(u)$. Dove $p(u) = f(x(u))$, con $x(u)$ il punto di minimo del problema vincolato, in corrispondenza del parametro u. In particolare: $\nabla p(0) = -\lambda^\star$
C'è il significato, importante, dei moltiplicatori come "prezzi ombra". Rapporto fra "guadagno" e la variazione di quantità disponibile di una risorsa vale λ.
Sensitività e interpretazione dei moltiplicatori di Kuhn-Tucker. Anche qui per i dettagli vedi Bertsekas. Se consideriamo il problema di minimizzare $f(x)$ con i vincoli $h(x) = u$ e $g(x) \le v$, , abbiamo che $\nabla_u p(u, v) = -\lambda(u, v)$ e $\nabla_v p(u, v) = -\mu(u, v)$. Qui $p(u, v) = f(x(u, v))$, con $x(u)$ il punto di minimo del problema vincolato, in corrispondenza del parametro (u, v). In particolare: $\nabla_u p(0, 0) = -\lambda^\star$ e $\nabla_v p(0, 0) = -\mu^\star$
Notare, a proposito di "prezzi ombra" che i moltiplicatori dei vincoli di disuguaglianza non attivi valgono 0. Infatti, all'ottimo sto usando quella risorsa al di sotto della disponibilità che ne ho e quindi non sono interessato a pagare per averne di più.

A livello di intuizione, ho $g(x) \le v$. Quindi se v aumenta il vincolo si allarga. Allora il $\min$ diminuisce. Cioè $p(v)$ diminuisce. Quindi è ragionevole pensare che:
$\nabla_v p(u, v) \le 0$ se e solo se $-\mu_j^\star \le 0$ (se e solo se $\mu_j^\star \ge 0$)

Vediamo quattro esempi: due dedicati ai vincoli di uguaglianza (e moltiplicatori di Lagrange), due dedicati a vincoli di disuguaglianza (e moltiplicatori di Kuhn - Tucker), ottenuti dai precedenti semplicemente convertendo le uguaglianze in disuguaglianze.

Primo esempio.
max $x + y$
con vincolo:
$x^2 + y^2 = b$
Aggiungiamo la condizione $b > 0$ sul parametro (così almeno il vincolo descrive una circonferenza, di raggio positivo). Il problema è semplicissimo e si risolve "a occhio". Il punto di max (globale) è unico ed è $(x^\star(b), y^\star(b)) = \left(\frac{\sqrt{2b}}{2}, \frac{\sqrt{2b}}{2}\right)$. Il valore ottimo della funzione è

$f(x^\star(b), y^\star(b)) = \sqrt{2b}$. E, naturalmente:
$\frac{d}{db}f(x^\star(b), y^\star(b)) = \frac{1}{\sqrt{2b}}$.
Scriviamo il sistema datoci dal teorema sui moltiplicatori di Lagrange (si verifica agevolmente che le condizioni che permettono di applicarlo sono soddisfatte). La lagrangiana è $L(x, y, \lambda) = (x + y) - \lambda\left(x^2 + y^2 - b\right)$.

$$\begin{cases} 1 - 2\lambda x = 0 \\ 1 - 2\lambda y = 0 \\ x^2 + y^2 = b \end{cases}$$

Da cui:

$$\begin{cases} x = \frac{1}{2\lambda} \\ y = \frac{1}{2\lambda} \\ \frac{2}{4\lambda^2} = b \end{cases}$$

Ovviamente i valori di $(x^\star(b), y^\star(b))$ sono gli stessi già provati. Ci interessa invece $\lambda^\star$. Da $\frac{2}{4\lambda^2} = b$, otteniamo $\lambda^2 = \frac{1}{2b}$. Da cui $\lambda = \pm\sqrt{\frac{1}{2b}}$, e il valore di λ che ci interessa è $\lambda^\star = \sqrt{\frac{1}{2b}}$ (l'altro valore corrisponde al punto di minimo assoluto).

Abbiamo quindi: $\lambda^\star = \frac{d}{db}f(x^\star(b), y^\star(b))$, come previsto dalla teoria.

Secondo esempio.
max $-(x-1)^2 - y^2$
con vincolo:
$x + y = b$
Anche questo problema si risolve "a occhio". Il punto di max (globale) è unico ed sta all'incrocio fra le rette $x + y = b$ e $y = x - 1$. E quindi abbiamo: $(x^\star(b), y^\star(b)) = \left(\frac{b+1}{2}, \frac{b-1}{2}\right)$. Il valore ottimo della funzione è $f(x^\star(b), y^\star(b)) = -\frac{(b-1)^2}{2}$. E, naturalmente:
$\frac{d}{db}f(x^\star(b), y^\star(b)) = 1 - b$
Scriviamo il sistema datoci dal teorema sui moltiplicatori di Lagrange (si verifica agevolmente anche in questo caso che le condizioni che permettono di applicarlo sono soddisfatte). La lagrangiana è $L(x, y, \lambda) = -(x-1)^2 - y^2 - \lambda(x + y - b)$.

$$\begin{cases} -2(x-1) - \lambda = 0 \\ -2y\lambda = 0 \\ x + y = b \end{cases}$$

Da cui:

$$\begin{cases} x = -\frac{\lambda - 2}{2} \\ y = -\frac{\lambda}{2} \\ \lambda = 1 - b \end{cases}$$

Ovviamente i valori di $(x^\star(b), y^\star(b))$ sono gli stessi già provati. Invece $\lambda^\star = 1 - b$.
Abbiamo quindi: $\lambda^\star = \frac{d}{db}f(x^\star(b), y^\star(b))$, come previsto dalla teoria.
Si noti che, per $b = 1$, abbiamo che la derivata è nulla. Non a caso per $b = 1$ il punto di massimo vincolato coincide col punto di massimo libero.
Terzo esempio (come il primo, tranne che ora abbiamo un vincolo di disuguaglianza).
max $x + y$
con vincolo:
$x^2 + y^2 \leq b$
Sempre con la condizione $b > 0$ sul parametro. Il punto di max (globale) è unico ed è $(x^\star(b), y^\star(b)) = \left(\frac{\sqrt{2b}}{2}, \frac{\sqrt{2b}}{2}\right)$, esattamente come succedeva col vincolo di uguaglianza. Il valore ottimo della funzione è $f(x^\star(b), y^\star(b)) = \sqrt{2b}$. E, naturalmente:
$\frac{d}{db}f(x^\star(b), y^\star(b)) = \frac{1}{\sqrt{2b}}$.
Scriviamo il sistema datoci dal teorema sui moltiplicatori di Kuhn - Tucker (si verifica agevolmente che le condizioni che permettono di applicarlo sono soddisfatte). La lagrangiana è identica al caso dei vincoli di uguaglianza, solo decido di chiamare μ il moltiplicatore solo per rendere più chiare le analogie e differenze con ciò che avviene con Lagrange: $L(x, y, \lambda) = (x + y) - \mu\left(x^2 + y^2 - b\right)$.

$$\begin{cases} \frac{\partial L}{\partial x} = 0 \\ \frac{\partial L}{\partial y} = 0 \\ \frac{\partial L}{\partial \mu} \le 0 \\ \mu \ge 0 \\ \mu \frac{\partial L}{\partial \mu} = 0 \end{cases}$$

Scindiamo il sistema dato in due, considerando il caso (1), quando $\frac{\partial L}{\partial \mu} \le 0$, e il caso (2), quando $\mu = 0$. Si noti che i due casi si possono parzialmente "sovrapporre", in quanto non possiamo escludere che sia abbia sia \frac{\partial L}{\partial \mu} = 0 che $\mu = 0$.

Caso (1):

$$\begin{cases} \frac{\partial L}{\partial x} = 0 \\ \frac{\partial L}{\partial y} = 0 \\ \frac{\partial L}{\partial \mu} = 0 \\ \mu \ge 0 \end{cases}$$

che ci dà: $\begin{cases} 1 - 2\mu x = 0 \\ 1 - 2\mu y = 0 \\ x^2 + y^2 = b \end{cases}$

Assolutamente identico a quello trovato nel primo esempio. In particolare, otteniamo: $\mu = \frac{1}{\sqrt{2b}}$.

Caso (2):

$$\begin{cases} \frac{\partial L}{\partial x} = 0 \\ \frac{\partial L}{\partial y} = 0 \\ \frac{\partial L}{\partial \mu} \le 0 \\ \mu = 0 \end{cases}$$

che ci dà, nella prima equazione: $1 = 0$, con poche speranze di poterla soddisfare...

Pertanto, questo terzo esempio ci dà risultati identici al primo. Fatto non sorprendente: il vincolo è sempre attivo, come si vede facilmente.

Quarto esempio. Anche qui, semplicemente convertiamo in vincolo di disuguaglianza il vincolo di uguaglianza che avevamo nel secondo esempio.
max $-(x-1)^2 - y^2$
con vincolo:
$x + y \le b$
Qui, se $b < 1$, il punto di max (globale) è unico ed sta allo incrocio fra le rette $x + y = b$ e $y = x - 1$. E quindi abbiamo:
$(x^\star(b), y^\star(b)) = \left(\frac{b+1}{2}, \frac{b-1}{2}\right)$. Il valore ottimo della funzione è $f(x^\star(b), y^\star(b)) = -\frac{(b-1)^2}{2}$. E, naturalmente:
$\frac{d}{db} f(x^\star(b), y^\star(b)) = 1 - b$
Quindi, per $b < 1$, si riproduce la stessa situazione che si aveva nel secondo esempio.
Se abbiamo $b \ge 1$, il punto di massimo globale vincolato coincide col punto di massimo globale libero (che è, ovviamente, $(1, 0)$). E quindi $f(x^\star(b), y^\star(b)) = f(1, 0) = 0$. Come si vede, il valore ottimo non dipende da b e quindi $\frac{d}{db} f(x^\star(b), y^\star(b)) = 0$
Scriviamo il sistema datoci dal teorema sui moltiplicatori di Kuhn - Tucker (le condizioni che permettono di applicarlo sono soddisfatte). La lagrangiana è $L(x, y, \mu) = -(x-1)^2 - y^2 - \mu(x + y - b)$.
Anche qui, usiamo due casi:
Caso (1):

$$\begin{cases} \frac{\partial L}{\partial x} = 0 \\ \frac{\partial L}{\partial y} = 0 \\ \frac{\partial L}{\partial \mu} = 0 \\ \mu \geq 0 \end{cases}$$

che ci dà: $\begin{cases} -2(x-1) - \mu = 0 \\ -2y\mu = 0 \\ x + y = b \\ \mu \geq 0 \end{cases}$

Da cui:

$$\begin{cases} x = -\frac{\mu - 2}{2} \\ y = -\frac{\mu}{2} \\ \mu = 1 - b \\ \mu \geq 0 \end{cases}$$

Ovviamente i valori di $(x^\star(b), y^\star(b))$ sono gli stessi già provati. Invece $\mu^\star = 1 - b$. Ma attenzione, che qui abbiamo anche la condizione $\mu \geq 0$. Quindi, ad esempio, abbiamo che $\mu^\star = 1 - b$ (e di conseguenza anche i valori delle $(x^\star, y^\star)$) solo per $1 - b \geq 0$.

Caso (2):

$$\begin{cases} \frac{\partial L}{\partial x} = 0 \\ \frac{\partial L}{\partial y} = 0 \\ \frac{\partial L}{\partial \mu} \leq 0 \\ \mu = 0 \end{cases}$$

che ci dà, dalle prime due equazioni: $x = 1$ e $y = 0$. Poi, dalla disequazione $\left(\frac{\partial L}{\partial \mu} \leq 0\right)$, ovvero $x + y \leq b$, ci dice che $b \geq 1$. E, naturalmente, $\mu = 0$.

Per la sensitività in PL, osserviamo che il teorema di dualità forte ci fornisce anche una informazione ulteriore. Se abbiamo una soluzione ottimale $x^\star = \begin{bmatrix} x_B^\star \\ 0 \end{bmatrix}$ relativa ad una base B, allora $y^\star = c_B B^{-1}$ è una soluzione ottimale del problema duale (questa è la strada usata da Tadei e Della Croce per dimostrare il teorema di dualità forte).

Da questa relazione e dalle condizioni di ortogonalità, scritte per un problema primale in forma standard, si può stimare la variazione della soluzione ottima e del valore ottimo al variare dei dati del problema di PL.

In particolare, se passiamo da b a $b + \Delta b$, la variazione del valore ottimo è:

$$\Delta z = c_B B^{-1}(b + \Delta b) - c_B B^{-1} b = c_B B^{-1} \Delta b = y^\star \Delta b = \sum_{i=1}^{m} y_i^\star \Delta b_i$$

Significato di $y_i^\star$ come prezzi ombra, nel problema di produzione.

Per la sensitività in PL, ricordo che abbiamo visto che: passando da b a $b + \Delta b$, la variazione del valore ottimo è:

$$\Delta z = c_B B^{-1}(b + \Delta b) - c_B B^{-1} b = c_B B^{-1} \Delta b = y^\star \Delta b = \sum_{i=1}^{m} y_i^\star \Delta b_i$$

Visti questi risultati nell'esempio solito (tratto da Fischetti).

Lu 29 marzo 2010, 2h-34h:
Introduzione. Decisioni in condizioni di: certezza, rischio, incertezza, ignoranza.
Decisioni in condizioni di certezza: esempio.
Modello: (X, E, h).
Preferenze su E e preordini totali: $\sqsupseteq$. Aspetti formali, preferenze "strette" ($\sqsupset$) e indifferenza ($\sim$).
Funzioni di utilità.
Modello ridotto.
Utilità ordinale (esempio: scala di Mohs; il concetto di temperatura).
Transitività: interpretazione, obiezioni, significato. Gli "atomi" di cioccolato. La "money pump".
DOCUMENTI:
Appunti (Moretti e Patrone) su decisioni in condizioni di certezza.

Gi 8 aprile 2010, 3h-37h:
Esistenza di funzione di utilità che rappresenta un preordine totale.
Una funzione di utilità (ordinale) è determinata a meno di una trasformazione strettamente crescente.
VAN. Altri esempi di criteri di scelta tra investimenti.
Problema del consumatore.
Introduzione alle decisioni in condizioni di rischio.
Stati di natura e loro probabilità.
DOCUMENTI:

Quadro concettuale.
Decisioni in condizioni di (certezza e di) rischio.
Appunti di Moretti: esempi di decisioni in condizioni di certezza (VAN, TIR, scelta del consumatore).

Lu 12 aprile 2010, 2h-39h:

$$\begin{pmatrix} X-S & | & s_1 & \dots & s_n \\ x_1 & | & \dots & \dots & \dots \\ \dots & | & \dots & \dots & \dots \\ x_m & | & \dots & \dots & \dots \end{pmatrix}$$

NB: S insieme finito.
Ad alternativa $\overline{x}$ assegno utilità attesa: $\sum_{s \in S} p(s)u(h(\overline{x}, s))$.
Preferenze su $K = \{k : S \to E\}$.
Avendo p data, "oggettiva" su S, ad una alternativa $\overline{x}$ resta associata una probabilità π su E, che ad $e \in E$ assegna probabilità $\pi(e) = \sum_{s: h(\overline{x}, s) = e} p(s)$.
Preferenze sulle lotterie (probabilità "semplici" su E, ovvero $\pi(e) = 0$ tranne che per un numero finito di elementi di E).
Tre condizioni su $\succeq$: preordine totale, indipendenza, condizione archimedea.
Allora esiste $u : E \to \mathbb{R}$:
$\pi_1 \succeq \pi_2 \Leftrightarrow \sum_{e \in E} \pi_1(e)u(e) \geq \sum_{e \in E} \pi_2(e)u(e)$
$U(\pi_1) = \sum_{e \in E} \pi_1(e)u(e)$ è la utilità attesa.
La u si dice utilità di vNM ed è definita a meno di trasformazioni affini strettamente crescenti ($v = au + b$ con $a > 0$ e $b \in \mathbb{R}$).
DOCUMENTI:
.
NOTA:
Ho usato il simbolo $\succeq$ anziché $\succcurlyeq$ per problemi di rendering (tra HTML, Unicode, MathML e chi più ne ha più ne metta...)

Lu 19 aprile 2010, 2h-41h:
Avversione al rischio: caso in cui $E \subseteq \mathbb{R}$. Confronto fra l'utilità attesa di una lotteria a esiti monetari e l'utilità attribuita all'esito certo pari al guadagno atteso.
Equivalente certo.
Indice di avversione al rischio di Arrow- Pratt. Le funzioni "CARA" ($u(x) = Ae^{-cx} + B$).
L'essenza del contratto assicurativo.
Scommesse: l'esempio del derby. Divergenza nella stima della probabilità.

Cenno. Il modello di Markovitz di scelta di portafoglio come "caso particolare" di decisioni in condizioni di rischio: utilità (di vNM) quadratica. Le determinanti della scelta sono il valore atteso e la varianza della lotteria.

Cenno. Dominanza stocastica del primo e del secondo ordine: criteri sufficienti di preferenza tra lotterie, sotto ipotesi abbastanza generali sulle proprietà della funzione di utilità di vNM del decisore.

DOCUMENTI:
Sul legame tra dominanza stocastica e utilità attesa, vedasi ad esempio gli appunti di Menoncin; qui copia in locale.

Gi 22 aprile 2010, 3h-44h:
Decisioni in condizioni di incertezza. Stessa formula, ma diversa interpretazione delle probabilità attribuite agli stati di natura.
Costruzione assiomatica a partire da $\sqsupseteq$ su $K = \{k : S \to E\}$. (Problemi: S insieme infinito).
Sure-thing principle: stesse ragioni della condizione di indipendenza.
Da $\sqsupseteq$ su $K = \{k : S \to E\}$ si ottiene $u : E \to \mathbb{R}$ e p, probabilità (soggettiva) su S.
Scambio libero (non forzoso): differenze in utilità o in probabilità.
Scommesse.
Speculazione $\neq$ investimento.
"No trade theorem": perché quello lì vuole vendere quelle azioni a un prezzo così basso?

Adverse selection, moral hazard.

Decisioni in condizioni di ignoranza (detta anche "completa incertezza" o "stretta incertezza"): il principio di maxmin di Wald (ma ce ne sono altri). Definisco $\varphi(x) = \min\left\{u\left(h(x, s_1)\right), \dots, u\left(h(x, s_n)\right)\right\}$, e scelgo $\overline{x}$ in X che massimizza φ.

Qualche assonanza fra maxmin e principio di precauzione.

Ottimizzazione vettoriale (multi-criterio).
Un decisore, più criteri.
Più decisori, ciascuno con un criterio.

$f : A \to B$. Visto caso $A = \mathbb{R}$. Ordine totale su B. Ma qui abbiamo $f : A \to \mathbb{R}^2$ (tratterò solo il caso $\mathbb{R}^2$, non cambia nulla se si passa ad $\mathbb{R}^n$).
Su $\mathbb{R}^2$ ordine totale (lessicografico). Ma non lo useremo (poco compatibile con la topologia di $\mathbb{R}^2$).
Usiamo ordine parziale: $\leq$
Ma dobbiamo passare da max a massimale (rara l'esistenza di max). Per esempio un insieme finito non necessariamente ha max rispetto a ordine parziale.
Massimale: $\overline{y} \geq y$ per ogni y **confrontabile** con $\overline{y}$.
Normalmente, non unicità. Esempi.
DOCUMENTI:

Brevi note su decisioni in condizioni di incertezza.
Appunti di Moretti su decisioni in condizioni di stretta incertezza.
Appunti su ottimizzazione vettoriale.

Lu 26 aprile 2010, 2h-46h:
Strutture matematiche: discussione.

Esempio: $\subseteq$.

Altre relazioni su $\mathbb{R}^2$: $<$, $\ll$.
Pareto (stretto): $\bar{y} \geq y$ per ogni ... confrontabile...
Si può introdurre $<$ su $\mathbb{R}^2$ a partire da $\leq$ (ordine stretto).
E ottimo paretiano si può definire equivalentemente come: non esiste y t.c. $y > \bar{y}$.
Se usiamo $\gg$, possiamo def ottimo paretiano DEBOLE: non esiste y t.c. $y \gg \bar{y}$.

Massimale: esiste, sotto ipotesi sensate?
Sì, se insieme è finito
Ma anche E, s.i. chiuso e limitato (non vuoto) di $\mathbb{R}^2$ ha elementi massimali. Basta massimizzare $ax + by$ su E, con $a, b \geq 0$, $(a, b) \neq (0, 0)$, che ha certamente max per Weierstrass.
Come si dimostra? Scalarizzazione.
Vedremo che sotto opportune ipotesi (convessità, essenzialmente) la strada può essere percorsa a rovescio.
Esempio (inizio): spartizione di 100 euro. Preferenze "egoistiche".

Attenzione: pesi/scala dei valori di utilità. Non farsi trarre in inganno.

DOCUMENTI:
.

Gi 29 aprile 2010, 3h-49h:
Esempio (fine).
Criterio lessicografico, altra strada. Che non corrisponde a mettere un peso pari a zero.

Costi/benefici

AHP

ELECTRE

Aggregativo-compensatore

Borda
Vedasi anche il GP del Brasile di Formula 1, 2007: durante il secondo pit-stop, Räikkönen riesce a sorpassare Massa. In tal modo vince la gara e anche il campionato, davanti ad Hamilton (per un punto...)
Graziano Battistini viene fatto fermare per aspettare Gastone Nencini in modo che potesse prendere l'abbuono (Tour de France del 1960, credo fosse la decima tappa, Mont-de-Marsan - Pau, vinta da Roger Rivière).

DOCUMENTI:
Introduction to Decision Making Methods, note di János Fülöp. Disponibili anche in locale.

Un interessante manuale di riferimento sulla analisi multi-criterio. Disponibile anche in locale.

Metodo aggregativo-compensatore, confronto a coppie:
SUPPLEMENTO ORDINARIO N. 66/L
DECRETO DEL PRESIDENTE DELLA REPUBBLICA 21 dicembre 1999, n.554
Regolamento di attuazione della legge quadro in materia di lavori pubblici 11 febbraio 1994, n. 109, e successive modificazioni.
Dal sito web del Comune di Jesi: http://gazzette.comune.jesi.an.it/2000/98/allsuppl66.htm

Su Borda e Arrow si possono consultare le prime 6 pagine di:
Implementazione. Con una breve introduzione alle scelte sociali e con l'esempio di Re Salomone.

Gi 6 maggio 2010, 2h-51h:
Statistica descrittiva.

DOCUMENTI:
Per statistica descrittiva, probabilità e statistica inferenziale:
appunti di Cufaro Petroni. Disponibile anche in locale. Essenzialmente seguo questo testo a lezione.
appunti di Garetto. Disponibile anche in locale.
Rogantin M.P.: *Introduzione alla statistica con esempi sviluppati in linguaggio Minitab*, CLUT, Torino, 2000.
Vedasi anche gli appunti (sempre di Rogantin) disponibili nella pagina dedicata al Progetto Lauree Scientifiche.
Per chi sia interessato ad un primo approfondimento: Paolo Baldi, *calcolo delle probabilità e statistica*, McGraw-Hill, Milano, 1993.
Un bel sito: sia in italiano che in inglese.

Gi 13 maggio 2010, 3h-54h:
Probabilità:
Spazi probabilizzabili, algebra di eventi (σ-algebra).
Probabilità, additività (e σ-additività)
Probabilità condizionata e indipendenza

Indipendenza di n eventi
Teorema di Bayes, esempio di applicazione
Variabile aleatoria
Legge (distribuzione) di una variabile aleatoria
Un esempio di due variabili aleatorie diverse ma identicamente distribuite
Funzione di ripartizione di una variabile aleatoria

DOCUMENTI:
.

Gi 20 maggio 2010, 3h-57h:
Probabilità:
Indipendenza di due variabili aleatorie
Esempi di leggi importanti (Bernoulli, binomiale, Poisson, uniforme, gaussiana; chi quadro, Student)
Convergenza di variabili aleatorie
Legge dei grandi numeri e teorema limite centrale

Statistica inferenziale
Un esempio da Rogantin
DOCUMENTI:
.

Gi 27 maggio 2010, 3h-60h:
Statistica inferenziale:
Stima di parametri
Stimatori non distorti e stimatori consistenti
Varianza campionaria
Intervalli di fiducia
Test di ipotesi
DOCUMENTI:
Gli appunti sulla varianza campionaria
Due cose che vale la pena conoscere:
Paradosso di Simpson: note di Anna Torre
Un approfondimento di Pearl in merito: Simpson's Paradox: An Anatomy
E qualche problema col p-value: The Earth is round ($p < .05$)

Per chi fosse interessato:
DOCUMENTI:
Teorema di Arrow e applicazione al pentathlon:
Arrow's Theorem: The Paradox of Social Choice - A Case Study in the Philosophy of Economics di Alfred F. MacKay, Yale University Press, 1984.
Conferenze all'IRRE: parte 3: scelte sociali
Su QALY:
QALY maximisation and people's preferences: a methodological review of the literature. Disponibile anche in locale.
QALYs: The Basics. Disponibile anche in locale.

BIBLIOGRAFIA:
Una breve bibliografia di Ricerca Operativa [http://www.diptem.unige.it/patrone/Biblio_commentata_intro_RO.htm].

Una breve bibliografia di Teoria delle Decisioni.

Può valere la pena di consultare i seguenti appunti (di Marco Sciandrone), disponibili in rete:
condizioni di ottimalità
proprietà generali di convergenza di algoritmi
algoritmi di ricerca unidimensionali
metodo del gradiente
metodo di Newton
Può essere interessante anche consultare la pagina del materiale didattico di Sciandrone.

Ultimo aggiornamento: 31 maggio 2010 (rispetto a versione del 29 maggio, aggiunto link a un documento, non essenziale, per la lezione del 10 novembre).

Ritorna alla home page [http://www.diptem.unige.it/patrone/default.htm] di Patrone