



Università degli Studi di Pisa
Dipartimento di Matematica
Corso di Laurea Magistrale in Matematica

Seminario di Teoria e Metodi dell'Ottimizzazione

Modelli finanziari di programmazione quadratica per l'ottimizzazione di portafoglio

Docente il PROFESSOR GIANCARLO BIGI

MARCO TARSIA

- mtarsia@student.dm.unipi.it -

Luglio 2017

Abstract

È ormai compresa e ben nota la fondamentale importanza dei modelli matematici di ottimizzazione per trattare efficientemente tutta una larga varietà di attività umane di notevole interesse, fra le quali per esempio quelle di natura finanziaria. In questo seminario lavoriamo su quella teoria rilevante e quindi su quei metodi risolutivi che ci permettano di analizzare appunto uno specifico tipo di problema di finanza matematica, quello di *strategia di portafoglio ottimo*. Con portafoglio intendiamo la rappresentazione dell'insieme di tutte le diverse attività finanziarie detenute da un dato investitore.

L'idea essenziale che sta alla base di un tale problema può essere presentata come segue: disponendo di una quantità limitata di risorse all'interno di un "certo" mercato, vorremmo selezionare quella od una combinazione d'investimenti diversificata abbastanza affinché il corrispondente portafoglio di valori risulti avere a rischio "basso" un rendimento atteso "non basso" per una prefissata data futura.

Una questione del genere è tanto interessante all'interno di una qualche situazione finanziaria concreta, ovviamente, quanto non banale all'interno di un qualche modello matematico realistico, poiché in generale per poter sperare verosimilmente in dei seri guadagni risulta necessario affrontare dei seri rischi, e per ciò nasce chiaramente un vero e problematico conflitto tra le due quantità rendimento atteso e rischio ad esso associato.

La modellizzazione che proponiamo e che studiamo in dettaglio di questo problema è quella data dal classico *modello di Ottimizzazione Media-Varianza* di Harry Markowitz¹ - o brevemente *modello MVO* di Markowitz - secondo cui il problema in questione di ottimizzazione di portafoglio è un particolare problema di *programmazione quadratica* - o brevemente di *QP* - e dunque un problema di ottimizzazione convessa. Formalmente, il problema equivale infatti a quello di minimizzare un'opportuna funzione scalare genuinamente quadratica che sia soggetta ad opportuni vincoli funzionali lineari, ed il tutto a dimensione tipicamente "alta" ma comunque finita.

Il riferimento bibliografico principale consiste certamente dei capitoli settimo ed ottavo di (2).

1 Programmazione Quadratica

1.1 Introduzione

Dato $n \in \mathbb{N} \setminus \{0\}$, un problema di ottimizzazione su \mathbb{R}^n è un problema di minimo vincolato su \mathbb{R}^n : data una qualsiasi regione ammissibile $D \subseteq \mathbb{R}^n$ infinita e data una qualsiasi funzione obiettivo $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ scalare, "trovare" il valor minimo di f su D , ossia il numero $\min_D f := \min\{f(x) : x \in D\}$. Più precisamente, ciò significa determinare $\inf_D f := \inf\{f(x) : x \in D\}$ e stabilire quindi se esso coincide davvero col valor minimo di f su D , ovvero se esiste almeno un $x^* \in D$ tale per cui risulti $f(x^*) \leq f(x)$ per ogni altro punto ammissibile $x \in D$, ed in tal caso individuare il sottoinsieme dei punti di minimo globale di f su D , ossia $\operatorname{argmin}_D f := \{x^* \in D : \forall x \in D, f(x^*) \leq f(x)\}$. Riferiamoci ad un tale problema col simbolo (\mathcal{P}) .

¹**Harry M. Markowitz**: rinomato economista statunitense nato a Chicago nell'agosto del 1927 e vincitore del premio Nobel per l'economia nel 1990, insieme a Merton H. Miller e William F. Sharpe, «per i contributi pionieristici nell'ambito dell'economia finanziaria». Il contributo più notevole da parte di H. Markowitz corrisponde proprio a tutto il suo lavoro svolto sulla diversificazione ottima degli investimenti, lavoro che ha origine in un suo articolo pubblicato nell'anno 1952 (per l'esattezza, in *Portfolio Selection* del Journal of Finance 7). **Nota**: a riguardo invece degli altri, vari scienziati che citeremo nel corso del seminario, per lo più matematici, non saremo altrettanto dettagliati sulle relative note biografiche.

Rimarchiamo a proposito che il saper risolvere ogni problema tipo (\mathcal{P}) è equivalente al saper risolvere ogni problema di massimo vincolato su \mathbb{R}^n : data una qualsiasi regione ammissibile $\tilde{D} \subseteq \mathbb{R}^n$ infinita e data una qualsiasi funzione obiettivo $\tilde{f}(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ scalare, determinare $\sup_{\tilde{D}} \tilde{f} := \sup\{\tilde{f}(x) : x \in \tilde{D}\}$, stabilire se esso è il valor massimo di \tilde{f} su \tilde{D} ed in caso affermativo individuare gli $x^* \in \tilde{D}$ tali che sia $\tilde{f}(x^*) \geq \tilde{f}(x)$ per ogni $x \in \tilde{D}$. Riferiamoci ad un tale problema col simbolo $(\tilde{\mathcal{P}})$.

Come poter approssimare e risolvere un problema di ottimizzazione su \mathbb{R}^n , che sia nel formalismo di (\mathcal{P}) o in quello di $(\tilde{\mathcal{P}})$, è idealmente scovare quelle caratteristiche analitico-geometriche della regione ammissibile e della funzione obiettivo su di essa che siano sufficienti a garantire a priori l'esistenza di una soluzione del problema in virtù di una qualche teoria astratta ben conosciuta, per poi utilizzare infine dei metodi quanto più efficaci possibile per ricavarla esplicitamente con una approssimazione buona a piacimento, se non con esattezza.

Teniamo presente che quanto dovrebbe esser più apprezzabile di un certo algoritmo risolutivo è che sia a tempo polinomiale, ovvero il cui tempo di lavoro possa esser limitato superiormente da una ben precisa funzione polinomiale della dimensione dei dati d'input, e ciò per ogni esempio specifico di tutta la classe di problemi per la quale è stato ideato. La dimensione di questi problemi può esser intesa come uguale allo spazio richiesto per scrivere tutti i propri dati di input in codice binario.

La teoria ed i metodi dell'ottimizzazione su \mathbb{R}^n che in questo senso c'interessano per i nostri scopi sono senz'altro di ottimizzazione convessa non-lineare e liscia per un problema tipo (\mathcal{P}) , rispettivamente di ottimizzazione concava non-lineare e liscia per un problema tipo $(\tilde{\mathcal{P}})$; ottimizzazione cioè di una funzione obiettivo convessa, rispettivamente concava, non-lineare e di classe \mathcal{C}^2 , o addirittura di classe \mathcal{C}^∞ - quindi un campo scalare -, su di una regione ammissibile convessa, chiusa e a parte interna non vuota, o addirittura di tipo poliedrale, di \mathbb{R}^n .

Infatti, la speciale classe di problemi d'ottimizzazione su \mathbb{R}^n alla quale ci restringiamo è quella dei problemi di *programmazione quadratica* su \mathbb{R}^n , di acronimo inglese *QP*, e chiamati anche di *ottimizzazione quadratica*: scegliendo ad esempio il formalismo di (\mathcal{P}) , la QP su \mathbb{R}^n si riferisce al problema di minimizzare una funzione obiettivo $f(x)$ quadratica su di una regione ammissibile D la quale sia un poliedro di \mathbb{R}^n , caratterizzabile quindi per mezzo di alcune equazioni e disequazioni lineari mutuamente coerenti. Possiamo rappresentare matematicamente un tale problema di QP nel modo seguente:

$$\min\left\{\frac{1}{2}x^T Qx + c^T x + a : \forall i = 1, \dots, p \wedge \forall j = 1, \dots, q, g_i(x) = 0 \wedge h_j(x) \geq 0\right\},$$

dove $x \in \mathbb{R}^n$ funge da solita variabile indipendente, mentre son tutti da intendersi come dati $Q \in \mathbb{R}^{n^2}$, $c \in \mathbb{R}^n$, $a \in \mathbb{R}$, p e $q \in \mathbb{N} \setminus \{0\}$, e $g_i(x), h_j(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ funzioni lineari per ogni $i = 1, \dots, p$ e $j = 1, \dots, q$.

Resta indubbiamente intuitivo che quanto più i numeri n , p e q sono elevati, tanto più la risoluzione di questo problema diventa delicata e laboriosa. Tuttavia sono precisamente la presenza ed il numero dei vincoli esprimibili in disequazioni fra quei fattori che contribuiscono maggiormente alla difficoltà di risolvere un qualsiasi problema di ottimizzazione, e questo lo comprenderemo personalmente più avanti.

La forma più comune del generico problema di QP, quella che può esser considerata *standard*, non coincide in verità esattamente con la precedente, sebbene comunque risulti ad essa equivalente, ed è piuttosto descrivibile così:

$$\min\left\{\frac{1}{2}x^T Qx + c^T x + a : Ax = b, x \geq 0\right\},$$

dove $x \in \mathbb{R}^n$ resta la variabile indipendente, dove ancora $Q \in \mathbb{R}^{n^2}$, $c \in \mathbb{R}^n$ e $a \in \mathbb{R}$, e dove invece sono dati pure $m \in \mathbb{N} \setminus \{0\}$, $A \in \mathbb{R}^{m \times n}$ ed infine $b \in \mathbb{R}^m$.

Qua sopra, per ogni $N \in \mathbb{N} \setminus \{0\}$ ed ogni $v \in \mathbb{R}^N$, stiamo considerando v come un vettore colonna di N componenti scalari ordinate v_1, \dots, v_N , o cioè come una matrice reale $N \times 1$, ed inoltre la notazione T corrisponde all'operazione matriciale di trasposizione, per cui v^T coincide col vettore trasposto di v stesso ed è quindi una matrice reale $1 \times N$. In particolare abbiamo che $v = (v_1, \dots, v_N)^T$, o cioè che $v^T = (v_1, \dots, v_N)$, e dunque che, per ogni $x \in \mathbb{R}^n$, valgano le due identità $c^T x = \langle c; x \rangle_2 = \sum_{k=1}^n c_k x_k$ e $x^T Qx = \langle Qx; x \rangle_2 = \sum_{k,k'=1}^n Q_{k,k'} x_k x_{k'}$. Invece con la notazione $v \geq 0$ intendiamo semplicemente che $v \in \mathbb{R}_+^N$, dove $\mathbb{R}_+ = [0, \infty)$, o esplicitamente che $v_i \geq 0$ per ogni $i = 1, \dots, N$. Così, considerato pure $w = (w_1, \dots, w_N)^T \in \mathbb{R}^N$ qualsiasi, con $v \geq w$ intendiamo che $v - w \geq 0$.

Pertanto, da una parte, la regione ammissibile D è

$$D = \{x \in \mathbb{R}^n : \forall i = 1, \dots, p \wedge \forall j = 1, \dots, q, g_i(x) = 0 \wedge h_j(x) \geq 0\},$$

oppure in forma standard

$$D = \{x \in \mathbb{R}^n : Ax = b \wedge x \geq 0\},$$

e per essa stiamo tacitamente supponendo che tutti i vincoli funzionali dati dalle g_i e dalle h_j , che siano espressi in forma standard o no, siano non soltanto non-contraddittori tra loro, ma siano anzi tali per cui D risulti in definitiva un vero poliedro non vuoto di \mathbb{R}^n , ed in particolare un insieme convesso, chiuso e limitato, quindi compatto, oltre che a parte interna non vuota, di \mathbb{R}^n . D'altra parte, la funzione obiettivo $f(x)$ è

$$f(x) = \frac{1}{2}x^T Qx + c^T x + a = \left\langle \frac{1}{2}Qx + c; x \right\rangle_2 + a, \quad x \in \mathbb{R}^n,$$

ed invece per essa supponiamo innocuamente ed una volta per tutte che abbia $a = 0$ e soprattutto che la matrice Q dei coefficienti delle seconde potenze sia una matrice simmetrica, ossia di matrice trasposta $Q^T = Q$ stessa, ed avente di conseguenza autovalori tutti reali.

D'ora innanzi utilizzeremo quasi sempre la seguente comoda notazione per il nostro problema (\mathcal{P}) :

$$\begin{cases} \min_x \frac{1}{2}x^T Qx + c^T x \\ Ax = b \\ x \geq 0. \end{cases}$$

Constatiamo così che la funzione obiettivo f è largamente di classe C^∞ su tutto \mathbb{R}^n ed è in generale sì non-lineare, ma pure potenzialmente lineare, oppure quasi-lineare, notata la possibilità perfettamente lecita di avere $Q = 0$, oppure $Q \simeq 0$. Osservato poi che la funzione f ha gradiente lineare dato da $\nabla f(x) = \text{grad } f(x)^T = Qx + c$, $x \in \mathbb{R}^n$ - essendo $\nabla f(\bar{x})_i = \frac{\partial f}{\partial x_i}(x)|_{x=\bar{x}}$ per ogni $\bar{x} \in \mathbb{R}^n$ e $i = 1, \dots, n$ - e ha quindi matrice hessiana costante data da $\nabla^2 f(x) = H_f(x) \equiv Q$, $x \in \mathbb{R}^n$ - essendo $\nabla^2 f(\bar{x})_{i,j} = \frac{\partial^2 f}{\partial x_i \partial x_j}(x)|_{x=\bar{x}}$ per ogni $\bar{x} \in \mathbb{R}^n$ e $i, j = 1, \dots, n$ -, deduciamo in più che f è convessa su \mathbb{R}^n se e solo se la matrice Q è semi-definita positiva, ossia tale che risulti $\langle Qx; x \rangle_2 \geq 0$ per ogni $x \in \mathbb{R}^n$, ed avente di conseguenza autovalori tutti non-negativi. Inoltre, in una tale situazione, f risulta strettamente convessa su \mathbb{R}^n se e solo se risulta anzi fortemente convessa su \mathbb{R}^n , e ciò se e solo se la matrice Q risulta in realtà definita positiva, ossia con $\langle Qx; x \rangle_2 > 0$ per ogni $x \in \mathbb{R}^n \setminus \{0\}$, o equivalentemente invertibile, ed avente di conseguenza autovalori tutti positivi, dei quali il minimo può esser scelto come modulo di forte convessità di f .

Alla luce di tutto questo, e poggiandosi sempre sulla parte più basilare e nota della teoria di analisi convessa volta all'ottimizzazione, completiamo il tutto evidenziando che, se la matrice Q è semi-definita positiva oltre che simmetrica, o cioè corrisponde ad una matrice di covarianze, allora il problema (\mathcal{P}) ammette almeno una soluzione, nel senso che f ammette perlomeno un punto di minimo globale in D , ed in più o tale soluzione è unica oppure ve ne esistono infinite; se poi Q risulta definita positiva, cioè non-singolare, allora il problema (\mathcal{P}) ammette una ed una sola soluzione.

Supponiamo pertanto che la funzione $f(x) = \frac{1}{2}x^T Qx + c^T x$ abbia la Q sia simmetrica ché semi-definita positiva. Il vero problema a questo punto è allora quello di determinare un algoritmo soddisfacente per il calcolo di una soluzione di (\mathcal{P}) , ovvero di un punto di minimo globale $x^* \in D$ di f su D . Per quanto sia ovvio il fatto che l'ambito di lavoro debba restare quello della programmazione non-lineare - o brevemente NLP -, notiamo comunque che un'eventuale buona riuscita di un tale compito sarebbe una buona riuscita non solo del generico problema di programmazione quadratica, ma pure del generico problema della ben più familiare programmazione lineare - o brevemente LP -, potendo anche essere $Q = 0$.

Ebbene, quanto proporremo più avanti come algoritmo risolutivo soddisfacente di (\mathcal{P}) sarà *il metodo di punto interno* introdotto nel 1984 dal matematico indiano N. K. Karmarkar, e verrà anzitutto ispirato sia dalla LP ché dalla NLP sia per la rispettiva teoria rilevante - principalmente il concetto di dualità e le varie condizioni di ottimalità - ché per i rispettivi metodi risolutivi più efficienti - principalmente il noto metodo del simplesso di G. Dantzig per la LP (1947) ed il classico metodo di Newton per la NLP non vincolata. Dopo di ché, è naturale che nella pratica la dimensione di un dato problema da risolvere sarà solitamente grande a tal punto che risulti davvero indispensabile l'ausilio di un buon software risolutivo, uno dei tanti disponibili ad esempio in rete.

1.2 Richiami: Programmazione Lineare e Programmazione Non-Lineare

Cominciamo ricordando con una certa cura il concetto di dualità, e quindi le condizioni di ottimalità ed il metodo del simplesso di Dantzig, per un problema di LP, e facciamo questo principalmente per tre ragioni: la prima è dovuta al voler dedurre la dualità per un problema di QP procedendo comodamente in analogia; la seconda è dovuta al voler cominciare ad affrontare un primo, possibile problema di portafoglio ottimo, seppur veramente semplice; la terza, nonché la più importante, è dovuta al voler risolvere rigorosamente il problema di MVO di Markowitz anche in una sua variante davvero speciale, chiamata col nome di *modello di Konno-Yamazaki* o *modello MAD* (1991), secondo cui il problema può esser ricondotto appunto ad uno di LP anziché ad uno di QP vera e propria.

Continuiamo ricordando il metodo degli iperpiani tangenti di Newton volto ad un problema di ottimizzazione di NLP non vincolata - per una funzione obiettivo regolare -, e ricordando dopo il concetto di funzione lagrangiana e quindi le ben note condizioni di ottimalità di W. Karush, H. W. Kuhn e A. W. Tucker, o condizioni KKT (1951), per un problema di NLP invece vincolato - sempre per una funzione obiettivo regolare: infatti, come vedremo presto, le iterazioni del metodo di punto interno sono varianti piuttosto sofisticate del metodo di Newton, e sono basate per tutto il tempo sulle condizioni KKT formulate per un problema di QP.

1.2.1 Programmazione Lineare: il Metodo del Simplexso di Dantzig

Il generico problema di LP su \mathbb{R}^n in forma standard ha la seguente rappresentazione:

$$\begin{cases} \min_x c^T x \\ Ax = b \\ x \geq 0, \end{cases}$$

dove come prima son dati $n, m \in \mathbb{N} \setminus \{0\}$, $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$ e $b \in \mathbb{R}^m$ tali per cui la regione ammissibile risulti non vuota, e dunque un vero poliedro di \mathbb{R}^n . Sappiamo allora che questo problema ammette almeno una soluzione, e più precisamente che almeno un vertice della regione ammissibile è un punto di minimo globale per la funzione lineare obiettivo. È per per riconoscere e per individuare una tale soluzione che entrano in gioco dapprima la dualità e le condizioni di ottimalità, ed infine il metodo del simplesso.

La semplice idea che sta alla base della nozione di problema duale associato a questo nostro problema di partenza, il quale prende di conseguenza il nome di problema primario, è quella di limitare “ottimalmente” dal basso la funzione obiettivo sfruttando “ottimalmente” i vari vincoli lineari del problema, e trovare quindi un punto ammissibile che raggiunga tramite la funzione obiettivo stessa tale limite inferiore. Illustriamo ciò mediante un elementare esempio.

Esempio 1. Consideriamo il seguente problema di LP su $\mathbb{R}_{(x_1, x_2)}^2$:

$$\begin{cases} \min_{(x_1, x_2)} -x_1 - x_2 \\ 2x_1 + x_2 \leq 12 \\ x_1 + 2x_2 \leq 9 \\ x_1 \geq 0, x_2 \geq 0. \end{cases}$$

Grazie a due variabili ausiliarie x_3 e x_4 , possiamo riscriverlo come problema di LP su $\mathbb{R}_{(x_1, x_2, x_3, x_4)}^4$ in forma standard:

$$\begin{cases} \min_{(x_1, x_2, x_3, x_4)} -x_1 - x_2 \\ 2x_1 + x_2 + x_3 = 12 \\ x_1 + 2x_2 + x_4 = 9 \\ x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0. \end{cases}$$

I vincoli di questo problema risultano chiaramente del tutto coerenti fra loro, e danno direttamente che $-x_1 - x_2 \geq -2x_1 - x_2 \geq -2x_1 - x_2 - x_3 = -12$, o ancora meglio che $-x_1 - x_2 \geq -x_1 - x_2 - x_3 - x_4 = -9$. Invece, combinandoli linearmente in modo oculato, otteniamo anzi che $-x_1 - x_2 \geq -x_1 - x_2 - \frac{1}{3}x_3 - \frac{1}{3}x_4 = -\frac{1}{3}(2x_1 + x_2 + x_3) - \frac{1}{3}(x_1 + 2x_2 + x_4) = -\frac{1}{3}(12+9) = -7$, e dunque che la funzione obiettivo $(x_1, x_2, x_3, x_4) \mapsto -x_1 - x_2$ è sempre ≥ -7 . Ma adesso possiamo notare subito che ad esempio $(5, 2, 0, 0)$ è ammissibile con $(5, 2, 0, 0) \mapsto -7$, ed in conclusione $(5, 2, 0, 0)$ è una soluzione del problema.

La strategia generale è stata quella di trovare una combinazione lineare dei due vincoli, diciamo con scalari y_1 per il primo e y_2 per il secondo, tale per cui il coefficiente così combinato di ciascuna variabile fornisca un limite inferiore per la funzione obiettivo rispetto a quella stessa variabile: più precisamente, cerchiamo $y_1, y_2 \in \mathbb{R}$ tali che $12y_1 + 9y_2 = (2x_1 + x_2 + x_3)y_1 + (x_1 + 2x_2 + x_4)y_2 = (2y_1 + y_2)x_1 + (y_1 + 2y_2)x_2 + y_1x_3 + y_2x_4$ sia $\leq -x_1 - x_2$, ossia (per non-negatività)

$$\begin{cases} 2y_1 + y_2 \leq -1 \\ y_1 + 2y_2 \leq -1 \\ y_1 \leq 0, y_2 \leq 0, \end{cases}$$

o ancora meglio tali che risolvano il problema di LP su $\mathbb{R}_{(y_1, y_2)}^2$

$$\begin{cases} \max_{(y_1, y_2)} 12y_1 + 9y_2 \\ 2y_1 + y_2 \leq -1 \\ y_1 + 2y_2 \leq -1 \\ y_1 \leq 0, y_2 \leq 0. \end{cases}$$

Questo problema coincide esattamente col problema duale associato al problema primario iniziale.

In tutta generalità ed in modo del tutto analogo capiamo subito che il problema duale associato al nostro problema primario risulta essere il problema di LP su \mathbb{R}_y^m

$$\begin{cases} \max_y b^T y \\ A^T y \leq c, \end{cases}$$

ossia il problema di LP su $\mathbb{R}_{(y, s)}^{m+n}$ in forma standard

$$\begin{cases} \max_{(y, s)} b^T y \\ A^T y + s = c \\ s \geq 0. \end{cases}$$

Anzi, ancor più in generale ma sempre in analogia possiamo facilmente verificare che il problema di QP su \mathbb{R}_x^n

$$\begin{cases} \min_x \frac{1}{2}x^T Qx + c^T x \\ Ax = b \\ x \geq 0 \end{cases}$$

ammette come problema duale associato il problema di QP su $\mathbb{R}_{(x, y, s)}^{n+m+n}$ in forma standard

$$\begin{cases} \max_{x, y, s} b^T y - \frac{1}{2}x^T Qx \\ A^T y - Qx + s = c \\ x, s \geq 0, \end{cases}$$

per il quale possiamo notare soprattutto che la variabile $x \in \mathbb{R}^n$ del problema primario compare come variabile a tutti gli effetti pure nel problema duale, a meno che non sia in realtà $Q = 0$.

Comunque, le relazioni più fondamentali tra il problema primario di LP ed il problema duale ad esso associato sono ben note e le rammentiamo qui di seguito. In primo luogo, per ogni $x \in \mathbb{R}^n$ ammissibile per il problema primario - ovvero con $Ax = b$ e $x \geq 0$ - e per ogni $y \in \mathbb{R}^m$ ammissibile per il suo problema duale - ovvero con $A^T y \leq c$ - sussiste la disuguaglianza $c^T x \geq b^T y$, ed il numero non-negativo $c^T x - b^T y$ viene chiamato gap di dualità fra x e y . Dai vincoli di non-negatività $x \geq 0$ e $c - A^T y \geq 0$ deduciamo infatti che $0 \leq (c - A^T y)^T x = c^T x - y^T Ax$, dove infine $Ax = b$. In secondo luogo, il gap di dualità fra certi x e y è uguale a zero esattamente quando x è una soluzione del problema primario e contemporaneamente y è una soluzione del problema duale: se x e y sono tali che $c^T x = b^T y$, allora x è una soluzione del problema primario e y è una soluzione del problema duale; e viceversa se il problema primario, rispettivamente duale, ammette una soluzione x , rispettivamente y , allora il problema duale, rispettivamente primario, ha una soluzione y , rispettivamente x , tale per cui risulti $c^T x = b^T y$. Quest'ultima proposizione sarebbe invece di non immediata dimostrazione.

Pertanto, le condizioni di ottimalità relative al nostro problema primario di LP sono le seguenti: un $x \in \mathbb{R}^n$ è una soluzione del problema primario se e solo se x è ammissibile - $Ax = b$ e $x \geq 0$ - ed esiste un $y \in \mathbb{R}^m$ ammissibile per il problema duale - $A^T y \leq c$ - tale che non ci sia gap di dualità fra x e y stessi, nel senso che valga $c^T x = b^T y$.

Analizzando ulteriormente il gap di dualità fra x e y , vediamo subito che $c^T x = b^T y$ se e solo se $(c - A^T y)^T x = 0$, e per non-negatività ciò equivale ad avere che $s_i := (c - A^T y)_i = 0$ oppure $x_i = 0$ per ogni $i = 1, \dots, n$.

Possiamo dunque riscrivere le condizioni di ottimalità: un $x \in \mathbb{R}^n$ è una soluzione del primario se e solo se x è ammissibile - $Ax = b$ e $x \geq 0$ - ed esiste un $y \in \mathbb{R}^m$ ammissibile per il duale - $s := c - A^T y \geq 0$ - tale che valga $x_i s_i = 0$ per ogni $i = 1, \dots, n$.

Il problema primario viene risolto passando proprio per queste condizioni di ottimalità, utilizzando o il metodo del simplesso o il più generale metodo di punto interno. Qua descriviamo le iterazioni di simplesso, e lo facciamo attraverso un semplice esempio di selezione di portafoglio tenendo presente quel risultato chiave della teoria della LP secondo cui nel caso il problema di LP ammetta una soluzione allora a tale problema deve corrispondere una soluzione che coincida con un vertice della regione ammissibile, come anche già ricordato poc'anzi, per cui la ricerca di una soluzione ottima può diventare la ricerca di un vertice più speciale degli altri. Questo fatto è notevole in modo ovvio, giacché un poliedro di \mathbb{R}^n possiede sempre un numero finito di vertici. Il metodo del simplesso risolve un problema di LP muovendosi da un vertice ad un altro vertice adiacente, e confrontando quindi i corrispondenti valori della funzione obiettivo. Una difficoltà potrebbe senz'altro essere l'inizializzazione dell'algoritmo, ossia la scelta di un punto iniziale che sia un vertice del poliedro, ma in realtà quest'operazione viene sempre piuttosto facile.

Esempio 2. *Un gestore di un portafoglio obbligazionario - ovvero di bonds - ha tutte le intenzioni di allocare 100 000\$ in due differenti bonds: uno societario, o aziendale, ed uno statale. Il bond societario ha una resa del 4% - sul netto di quanto ci viene investito -, ha una scadenza di 3 anni e ha una A di rating - secondo giudizio di una certa azienda di rating -, la quale ai nostri fini di calcolo interpretiamo come rating numerico di 2. Il bond statale ha invece una resa del 3%, quindi inferiore, ma ha una scadenza di 4 anni e ha una Aaa di rating, quindi superiore, alla quale facciamo corrispondere ad un rating numerico di 1. In particolare il bond statale ha una più alta qualità stimata rispetto a quella del bond societario, perché è doppia. Il gestore del portafoglio vorrebbe dunque allocare i suoi fondi in un modo tale per cui il rating medio del portafoglio risultante - necessariamente una via di mezzo tra A e Aaa - non sia qualitativamente peggiore di un rating di Aa, o cioè non sia quantitativamente superiore ad un rating numerico di 1.5, ed inoltre che la scadenza media - necessariamente fra i 3 ed i 4 anni - sia al più di 3.6 anni. Immaginiamo che l'ammontare di denaro contante non investito nei due bonds venga mantenuto in un conto corrente che per semplicità viene supposto d'interesse trascurabile, o perlomeno che non contribuisca né al rating medio né alla scadenza media. In altre parole, stiamo supponendo che il denaro contante abbia qualità di rating massima 0 e scadenza di 0. Ebbene, tenendo presenti tutti i dati su resa, scadenza e rating, la questione è come allocare i fondi a disposizione fra questi due bonds per realizzare l'obiettivo di massimizzare il profitto da tale investimento, con la naturale speranza di possedere infine più ricchezza di quella iniziale. Lasciando che le due variabili x_1 e x_2 di \mathbb{R}_+ denotino sotto forma di migliaia di dollari la quantità di fondo allocata nel bond societario e nel bond statale rispettivamente, questo problema di portafoglio assume la seguente forma:*

$$\begin{cases} \max_{(x_1, x_2)} 4x_1 + 3x_2 \\ x_1 + x_2 \leq 100 \\ \frac{2x_1 + x_2}{100} \leq 1.5 \\ \frac{3x_1 + 4x_2}{100} \leq 3.6 \\ x_1, x_2 \geq 0, \end{cases}$$

dove notiamo in particolare che x_1, x_2 vanno prese in percentuale per il calcolo del rating medio e della scadenza media. La sua forma standard richiede tre variabili ausiliarie $x_3, x_4, x_5 \geq 0$ e tiene conto delle semplici assunzioni sul denaro contante:

$$\begin{cases} \max_{(x_1, x_2, x_3, x_4, x_5)} 4x_1 + 3x_2 \\ x_1 + x_2 + x_3 = 100 \\ 2x_1 + x_2 + x_4 = 150 \\ 3x_1 + 4x_2 + x_5 = 360 \\ x_1, x_2, x_3, x_4, x_5 \geq 0. \end{cases}$$

Quanto il metodo del simplesso prescrive di fare a questo punto è di passare da una prescelta soluzione ammissibile di base, di acronimo inglese BFS, ad un'altra BFS.

Ricordiamo di cosa si tratta per un generico problema di LP su \mathbb{R}_x^n della forma

$$\begin{cases} \max_x c^T x \\ Ax \leq b \\ x \geq 0, \end{cases}$$

il quale in forma standard diventa un problema di LP su $\mathbb{R}_{(x,x_s)}^{n+m}$

$$\begin{cases} \max_{(x,x_s)} c^T x \\ [A \quad I_m] \begin{bmatrix} x \\ x_s \end{bmatrix} = b \\ \begin{bmatrix} x \\ x_s \end{bmatrix} \geq 0, \end{cases}$$

dove con $I_m \in \mathbb{R}^{m \times m}$ indichiamo la matrice identica $m \times m$ e dove $x_s = (x_{n+1}, \dots, x_{n+m})^T \in \mathbb{R}^m$ è costituito dalle variabili ausiliarie necessarie e sufficienti per la riscrittura standard. Per il nostro esempio, in particolare, abbiamo $n = 2$ ed $m = 3$.

In linea di principio potrebbero esistere infinite soluzioni di questo problema. Già $x = 0$ unito $x_s = b$ soddisfano il vincolo di uguaglianza, ma certo non necessariamente quello di disuguaglianza $x_s \geq 0$ - in quanto potrebbe benissimo essere $b \not\geq 0$ - o tanto meno la proprietà di ottimalità. Ecco come possiamo ottenere una soluzione in generale: se B è una matrice $m \times m$ costruita in modo tale che le sue colonne siano combinazioni lineari di alcune delle colonne linearmente indipendenti della matrice $m \times (n+m)$ data da $[A \quad I_m]$, allora permutiamo le colonne di $[A \quad I_m]$ stessa per ottenere la matrice $[B \quad N]$, dove naturalmente $N = N(B) \in \mathbb{R}^{m \times n}$ è la matrice residuo di tale operazione, e di conseguenza permutiamo coerentemente pure gli elementi di $[x^T \quad x_s^T]^T$ ottenendo diciamo $[x_B^T \quad x_N^T]^T$, dove però $x_B \in \mathbb{R}^m$ mentre $x_N \in \mathbb{R}^n$, dimodoché l'uguaglianza diventi

$$[B \quad N] \begin{bmatrix} x_B \\ x_N \end{bmatrix} = Bx_B + Nx_N = b,$$

o anche $x_B + B^{-1}Nx_N = B^{-1}b$, e pertanto sono equivalenti le tre seguenti identità:

$$\begin{cases} [A \quad I_m] \begin{bmatrix} x \\ x_s \end{bmatrix} = b \\ Bx_B + Nx_N = b \\ x_B + B^{-1}Nx_N = B^{-1}b. \end{cases}$$

Una matrice B di questo tipo è chiamata matrice di base, ed in sua corrispondenza le variabili di x_B son dette variabili di base, mentre quelle di x_N sono le variabili libere o non di base. Osserviamo che una qualsiasi scelta delle n variabili libere di x_N forza automaticamente quelle di base ad essere tali che $x_B = B^{-1}b - B^{-1}Nx_N$, ed in particolare la scelta $x_N = 0$ dà $x_B = B^{-1}b$, il che generalizza quanto notato sopra per $B = I_m$. Una soluzione di base, di acronimo inglese BS, è proprio un vettore $[x_B^T \quad x_N^T]^T$ dove $x_N = 0$ e $x_B = B^{-1}b$, o cioè della forma

$$\begin{bmatrix} x_B \\ x_N \end{bmatrix} = \begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix},$$

per un'opportuna matrice di base B . Una BS è ammissibile, e dunque è una BFS, se e solo se vale $B^{-1}b \geq 0$, ed è facile rendersi conto che geometricamente una BFS è un vertice della regione ammissibile.

L'effetto netto sulla funzione obiettivo $Z := c^T x$ del possibile variare delle variabili non di base x_N è quanto ora più c'interessa, ed è argomentato in seguito: permutando coerentemente c in $[c_B^T \quad c_N^T]^T$, risulta ovviamente che

$$Z - [c_B^T \quad c_N^T] \begin{bmatrix} x_B \\ x_N \end{bmatrix} = 0,$$

ovvero che $Z - c_B^T x_B - c_N^T x_N = 0$, o equivalentemente che $Z - c_B^T (B^{-1}b - B^{-1}Nx_N) - c_N^T x_N = 0$, o anche

$$Z - (c_N^T - c_B^T B^{-1}N)x_N = c_B^T B^{-1}b,$$

dove non compaiono più le variabili di base x_B . Quest'ultima uguaglianza è la proprietà essenziale sulla quale si appoggia il metodo del simplesso per ciascuna delle sue iterazioni: intendendo infatti il vettore di quei coefficienti $c_N^T - c_B^T B^{-1}N$ della funzione obiettivo corrispondenti a x_N come il vettore dei costi ridotti - coefficienti non di base di costo c_N ridotti dagli effetti incrociati dei coefficienti di base di costo c_B -, l'idea è quella di muoversi da una BFS ad un'altra BFS, ossia da un vertice ad un altro adiacente - cambiando ogni volta la matrice di base B -, finché i costi ridotti diventano ≤ 0 , in quanto per una BFS vale sempre $Z = c_B^T B^{-1}b$ e così se appunto B è tale che $c_N^T - c_B^T B^{-1}N \leq 0$ allora chiaramente $Z \leq c_B^T B^{-1}b$ per ogni $x_N \geq 0$.

Per la precisione, questo metodo è chiamato del simplesso primario, in quanto esisterebbe pure un metodo detto del simplesso duale: cominciando da una qualsiasi BS non ammissibile e mantendendo via via i costi ridotti ≤ 0 per le BS successive, muoversi fra queste BS fino a trovarne una ammissibile.

Torniamo alla pratica applicando questo macchinario al problema iniziale di portafoglio dei due bonds. Per cominciare va scelto un vertice iniziale, e questo è davvero immediato in quanto $b = (100, 150, 360)^T \geq 0$, essendo addirittura > 0 : rimanendo infatti con $B = I_3$, dunque con $N = N(B) = A$, e scegliendo $x_N = x = (x_1, x_2)^T = 0$ e $x_B = (x_3, x_4, x_5)^T = B^{-1}b = b$, otteniamo già una BFS semplicemente perché appunto $b \geq 0$. Notiamo a proposito che $x_N = (0, 0) \mapsto 0$, per cui è ovvio che possiamo ottenere di meglio, e per farlo dobbiamo far variare le variabili non di base $x_N \geq 0$, o cioè scambiare almeno una fra x_1 e x_2 con una delle variabili di base x_3, x_4, x_5 . In effetti, or ora abbiamo fissato $x_1 = x_2 = 0$, e comunque deve sempre essere $x_1, x_2 \geq 0$, per cui cambiare una delle due significa incrementare l'altra, che diventerà così > 0 ; per mantenere l'ammissibilità data dal vincolo delle uguaglianze, questo richiederà quindi di aggiustare almeno una fra x_3, x_4, x_5 . Ma x_3, x_4, x_5 son tutte

> 0 per $x_1 = x_2 = 0$, per cui per esempio sarà possibile rendere $x_1 > 0$ senza violare alcun vincolo. Dato poi che nessuna fra x_3, x_4, x_5 compare nella funzione obiettivo, via via dobbiamo solo guardare il coefficiente costo ridotto di quella variabile non di base in x_N che vogliamo aumentare per vedere che effetto fa su $(x_1, x_2) \mapsto 4x_1 + 3x_2$. L'idea è modificare una sola fra x_1 e x_2 per volta - andando così di BFS in BFS davvero adiacenti -, ed essendo $4 > 3$ dovremmo arrivare prima al valor massimo della funzione obiettivo cominciando con incrementare prima x_1 . Visto ora che la base ha in ogni caso tante variabili ausiliarie quante sono le equazioni del vincolo - 3, in questo caso - x_1 deve necessariamente sostituire una fra x_3, x_4, x_5 . In quanto invece le variabili non di base hanno sempre valore zero per una BFS, dobbiamo determinare quanto incrementare x_1 in modo che una fra x_3, x_4, x_5 diventi zero e possa così esser scambiata con x_1 . Il vero problema resta sempre quello di mantenere tutte $x_3, x_4, x_5 \geq 0$, ma tuttavia per il nostro problema ciò si rivela semplice in quanto ciascuna di x_3, x_4, x_5 compare solo in una riga, ed in più x_1 ha sempre coefficienti positivi. Dunque, essendo $x_2 = 0$, i tre vincoli diventano

$$\begin{cases} x_1 + x_3 = 100 \\ 2x_1 + x_4 = 150 \\ 3x_1 + x_5 = 360, \end{cases}$$

da cui risultano immediate le implicazioni

$$\begin{cases} x_1 > 100 \Rightarrow x_3 < 0 \\ x_1 > 75 \Rightarrow x_4 < 0 \\ x_1 > 120 \Rightarrow x_5 < 0, \end{cases}$$

le quali corrispondono a tutte eventualità da evitare assolutamente: pertanto, dovrà essere $0 < x_1 \leq \min\{100, 75, 120\} = 75$. D'altra parte per $x_1 = 75$ è $x_4 = 0$ - mentre $x_3 = 25 > 0$ e $x_5 = 135 > 0$ - e per ciò può lasciare la base: così per $x_1 = 75$ le nuove variabili di base diventano $x_B = \{x_3, x_1, x_5\}$, mentre quelle non di base sono $x_N = (x_4, x_2) = 0$. Per questa nuova scelta abbiamo evidentemente

$$B = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 3 & 1 \end{bmatrix}, \quad x_b = B^{-1}b = \begin{bmatrix} 25 \\ 75 \\ 135 \end{bmatrix}, \quad e \quad N = \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ 1 & 4 \end{bmatrix}.$$

A questo punto è sicuramente più complicato migliorare, perché x_1 compare sia nella base sia nella funzione obiettivo. In effetti, se adesso decidessimo per esempio di aumentare x_2 , non sarebbe affatto detto che aumenterebbe anche il valore corrispondente della funzione obiettivo, poiché incrementare x_2 potrebbe benissimo implicare una diminuzione netta di x_1 . L'idea naturale è allora quella di determinare una qualche manipolazione che abbia l'effetto di far scomparire la x_1 dalla funzione obiettivo e anche dalle righe dove x_1 stessa non è propriamente una variabile di base. Ciò riusciamo ad ottenerlo sfruttando invece la riga che rende x_1 una variabile di base, ovvero la seconda: è $2x_1 + x_2 + x_4 = 150$ se e solo se $x_1 = 75 - \frac{1}{2}x_2 - \frac{1}{2}x_4$, da cui immediatamente la riscrittura dell'intero problema

$$\begin{cases} \max_{(x_1, x_2, x_3, x_4, x_5)} Z = \max_{(x_1, x_2, x_3, x_4, x_5)} 4x_1 + 3x_2 \\ Z - x_2 + 2x_4 = 300 \quad (\text{cioè } Z = x_2 - 2x_4 + 300) \\ \frac{1}{2}x_2 - \frac{1}{2}x_4 + x_3 = 25 \\ \frac{1}{2}x_2 + \frac{1}{2}x_4 + x_1 = 75 \\ \frac{5}{2}x_2 - \frac{3}{2}x_4 + x_5 = 135 \\ x_1, x_2, x_3, x_4, x_5 \geq 0. \end{cases}$$

Possiamo dunque ripartire come prima perché appunto la funzione obiettivo è nuovamente libera dalle variabili di base, le quali inoltre compaiono solo nelle righe dove sono di base e ciò con coefficiente addirittura uguale a 1. Osserviamo in particolare che incrementare x_2 di 1 aumenta Z di 1 e non di 3, mentre aumentare x_4 la fa addirittura diminuire. Incrementiamo quindi x_2 , mentre x_4 resta = 0: per $x_2 = 50$ è chiaramente $x_3 = 0$ - per $x_2 > 50$ viene $x_3 < 0$ che è inammissibile -, e dunque $x_2 = 50$ sostituisce x_3 nella base ed il sistema diventa di conseguenza

$$\begin{cases} Z = 350 \\ x_1 = 50 \\ x_2 = 50 \\ x_3 = x_4 = 0 \\ x_5 = 10, \end{cases}$$

e siamo arrivati alla fine in quanto non sono possibili altri movimenti. In conclusione, il valor massimo della funzione obiettivo è 350 e viene raggiunto in corrispondenza del vertice (50, 50).

L'interpretazione grafica di tutto questo è davvero elementare: abbiamo lavorato sul pentagono del quadrante di $\mathbb{R}^2_{(x_1, x_2)}$ con $x_1, x_2 \geq 0$ muovendoci di vertice in vertice adiacente ed in senso antiorario cominciando da (0, 0), proseguendo con (75, 0) incrementando x_1 - ma saremmo potuti passare a (0, 90) incrementando x_2 -, e terminando in (50, 50) incrementando x_2 .

Il metodo del simplesso risulta tipicamente molto veloce pure nel gestire problemi con migliaia di variabili e centinaia di vincoli, ed anzi con centinaia di migliaia di variabili e decine di migliaia di vincoli - come poi è abbastanza comune in finanza - nonostante che per problemi di dimensioni così elevate il numero d'iterazioni tenda a sua volta ad esser davvero grande. Tuttavia, sebbene appunto tale metodo dimostri performance soddisfacenti nella maggior parte dei problemi pratici, ha il deprecabile svantaggio teorico legato alla cosiddetta worst-case complexity: il totale del tempo di calcolo potrebbe potenzialmente crescere in modo esponenziale con la dimensione del problema, e per ciò il metodo del simplesso non è di certo a tempo polinomiale.

Teniamo in mente che se tutti i numeri in gioco restano limitati fra 10^{-6} e 10^6 allora potremmo verificare che una buona proxy per tale dimensione è data dal prodotto del numero delle variabili col numero dei vincoli. Una variabile proxy è una variabile non davvero rilevante di per sé, ma che lo diventa nel momento in cui viene designata a rimpiazzarne un'altra che sarebbe quella realmente interessante ma purtroppo non calcolabile con esattezza, ed inoltre una buona variabile proxy dev'essere perlomeno strettamente correlata a tale variabile.

Due algoritmi alternativi a quello del simplesso di Dantzig, primario o duale che sia, che risultano a tempo polinomiale sono quello dell'ellissoide di Yudin-Nemirovski (1979) e soprattutto quello di punto interno di Karmarkar, dei quali quest'ultimo si manifesta apprezzabilmente più veloce anzitutto nella pratica. Cosa accade col metodo di punto interno applicato ad un problema di LP è che i movimenti avvengono sempre all'interno della regione ammissibile, anziché sui vertici - uno dei quali andrà comunque raggiunto - con delle iterazioni sì piuttosto dispendiose ma più "libere" ed infatti in quantità tipicamente quasi indipendente dalle dimensioni del problema e spesso ≤ 50 .

Vogliamo comunque osservare quanto non sia affatto stupefacente una simile efficienza per i problemi di LP, pensando a quanto le funzioni lineari siano sempre così regolari e prevedibili. Ben diversa invece la situazione per la NLP: è tanto vero che le formulazioni più realistiche di tanti, importanti problemi di ottimizzazione non si adattano ad una struttura lineare e piuttosto ad una non-lineare, quanto però è vero che non esiste un singolo algoritmo privilegiato per il generico problema di NLP, poiché in pratica un metodo non-lineare può essere molto efficiente per un certo tipo di problema e tuttavia fallire miseramente per un altro.

Ripassiamo adesso la teoria più basilare sulla NLP per funzioni obiettivo regolari, mentre subito dopo studieremo il metodo di punto interno per la QP e pure qualcuna delle sue possibili varianti.

1.2.2 Programmazione Non-Lineare: il Metodo degli Iperpiani Tangenti di Newton

Il contesto è senz'altro quello della NLP senza vincoli: fissato $N \in \mathbb{N} \setminus \{0\}$ ed assegnata una funzione $F(x) : \mathbb{R}^N \rightarrow \mathbb{R}^N$ di classe almeno \mathcal{C}^1 su tutto \mathbb{R}^N - quindi un campo vettoriale - e di componenti $F_1(x), \dots, F_N(x) : \mathbb{R}^N \rightarrow \mathbb{R}$, dove $x = (x_1, \dots, x_N)^T$ come prima, il problema è quello di determinare uno zero di F , cioè un punto $x^* \in \mathbb{R}^N$ tale che $F(x^*) = 0$. Si tratta dunque di risolvere il seguente sistema non-lineare quadrato di N equazioni in N incognite:

$$\begin{cases} F_1(x_1, \dots, x_N) = 0 \\ \vdots \\ F_N(x_1, \dots, x_N) = 0. \end{cases}$$

Le iterazioni del metodo di Newton suppongono anzitutto che F ammetta veramente uno zero e richiedono prima di scegliere un qualsiasi $x^0 \in \mathbb{R}^N$ e poi di definire ricorsivamente $x^k \in \mathbb{R}^N$ per ogni $k = 1, 2, \dots$, ovvero $k \in \mathbb{N} \setminus \{0\}$, procedendo nel modo seguente: ordinatamente per ciascun $k = 0, 1, 2, \dots$, ovvero $k \in \mathbb{N}$, calcoliamo la matrice jacobiana $J_F(x^k)$ di F in x^k - tale che $J_F(\bar{x})_{i,j} = \frac{\partial f_i}{\partial x_j}(x)|_{x=\bar{x}}$ per ogni $\bar{x} \in \mathbb{R}^N$ e $i, j = 1, \dots, N$ - ed approssimiamo quindi al prim'ordine mediante lo sviluppo di Taylor la F proprio in x^k , ottenendo subito la relazione

$$F(x^k + \delta) \approx_{\|\delta\|_2 \downarrow 0} \widehat{F}^k(\delta) := F(x^k) + J_F(x^k)\delta.$$

Ora l'idea è quella di risolvere per ogni $k \rightarrow \infty$ l'equazione $\widehat{F}^k(\delta^*) = 0$ anziché quella di partenza $F(x^*) = 0$: più precisamente, per ogni $k \in \mathbb{N}$ e nel caso che $J_F(x^k)$ sia sempre non-singolare, visto che chiaramente $\widehat{F}^k(\delta_k^*) = 0$ se e solo se $\delta_k^* = -J_F(x^k)^{-1}F(x^k)$, la formula per il generico passo del metodo di Newton è

$$x^{k+1} := x^k + \delta_k^* = x^k - J_F(x^k)^{-1}F(x^k).$$

Quello che si può dimostrare a riguardo e che comunque è risaputo è la convergenza super-lineare del second'ordine del metodo, nel senso che se $(x^k)_{k \in \mathbb{N}}$ è una successione generata dal metodo allora esiste $x^* \in \mathbb{R}^N$ con $F(x^*) = 0$ tale per cui valga $x^k \rightarrow x^*$ per $k \rightarrow \infty$ con $\|x^{k+1} - x^*\|_2 = \mathcal{O}(\|x^k - x^*\|_2^2)$.

Una proprietà del genere è indubbiamente più che apprezzabile, ma certo il metodo presenta pure degli svantaggi rappresentati specialmente dalla necessità di calcolare tutte le derivate prime di ognuna delle componenti del campo e pure d'invertire matrici tipicamente di dimensioni elevate.

Comunque tutto questo resta interessante per l'ottimizzazione svincolata di funzioni sufficientemente regolari $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ in quanto in ogni caso la ricerca concernerebbe uno di quei punti $x^* \in \mathbb{R}^n$ tali che $\nabla f(x^*) = 0$, e per ciò se la f fosse di classe \mathcal{C}^2 su tutto \mathbb{R}^n dimodoché il campo vettoriale $F := \nabla f$ diventi di classe \mathcal{C}^1 su tutto \mathbb{R}^n allora risolvere $\nabla f(x^*) = 0$ sarebbe esattamente risolvere $F(x^*) = 0$, equazione per la quale potremmo applicare appunto il metodo di Newton - con $N = n$. In una situazione di questo tipo avremmo precisamente che $J_F = \nabla^2 f$ e dunque che la generica iterazione del metodo Newton assumerebbe la forma

$$x^{k+1} := x^k - \nabla^2 f(x^k)^{-1} \nabla f(x^k).$$

Osserviamo infine quanto risulti evidente che un qualche vincolo sugli x^k , come ad esempio potrebbe essere $x^k \geq 0$ per ogni $k \in \mathbb{N}$, possa mettere profondamente in crisi l'intero metodo.

1.2.3 Programmazione Non-Lineare: le Condizioni di Ottimalità di Karush-Kuhn-Tucker

Il contesto torna nuovamente ad essere quello della NLP in presenza di vincoli: concentriamoci infatti su di un generico problema di ottimizzazione di tipo (\mathcal{P}) su \mathbb{R}^n

$$\min\{f(x) : \forall i = 1, \dots, p \wedge \forall j = 1, \dots, q, g_i(x) = 0 \wedge h_j(x) \geq 0\},$$

per il quale supponiamo almeno che la funzione obiettivo $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ - non-lineare - e tutte le funzioni di vincolo $g_i(x), h_j(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ - non-lineari - siano di classe \mathcal{C}^1 su tutto \mathbb{R}^n . Affianchiamo dunque a tutte queste mappe differenziabili con continuità la funzione lagrangiana $\mathcal{L}(x, \lambda, \lambda') : \mathbb{R}^{n+p+q} \rightarrow \mathbb{R}$ relativa al problema, altra mappa a sua volta della medesima regolarità nonché lo fondamentale strumento per la risoluzione del problema stesso passando per le condizioni KKT - chiamate anche di qualificazione dei vincoli:

$$\mathcal{L}(x, \lambda, \lambda') = f(x) - \sum_{i=1}^p \lambda_i g_i(x) - \sum_{j=1}^q \lambda'_j h_j(x),$$

al variare di ogni $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ e di ogni moltiplicatore $\lambda = (\lambda_1, \dots, \lambda_p)^T \in \mathbb{R}^p, \lambda' = (\lambda'_1, \dots, \lambda'_q)^T \in \mathbb{R}^q$.

La proprietà essenziale a riguardo consiste nell'equivalenza tra il risolvere l'iniziale problema di minimo su \mathbb{R}^n vincolato alle $p+q$ relazioni funzionali date col minimizzare la lagrangiana associata su tutto il suo dominio \mathbb{R}^{n+p+q} , ossia in modo svincolato, come comunque risulta intuitivo già dalla costituzione stessa di tale lagrangiana.

Rammentiamo infatti i teoremi di Karush-Kuhn-Tucker sulle condizioni necessarie e sufficienti di ottimalità, ricordando dapprima che un punto $x^0 \in \mathbb{R}^n$ il quale sia ammissibile - ovvero con $g_i(x^0) = 0$ e $h_j(x^0) \geq 0$ per ogni i, j - viene detto regolare quando per ciascuno di quegli indici $j_0 \in \mathcal{J}_0 \subseteq \{1, \dots, q\}$ tali che valga esattamente $h_{j_0}(x^0) = 0$ accade che tutti i vettori $\nabla h_{j_0}(x^0)$ al variare proprio di $j_0 \in \mathcal{J}_0$ risultano linearmente indipendenti.

Allora le condizioni necessarie di ottimalità al prim'ordine di Karush-Kuhn-Tucker sono così descritte: se $x^* \in \mathbb{R}^n$ è un punto ammissibile - $g_i(x^*) = 0$ e $h_j(x^*) \geq 0$ per ogni i, j -, se è di minimo locale per la funzione obiettivo $f(x)$, e se inoltre è regolare, allora esistono $p+q$ moltiplicatori $\lambda_i, \lambda'_j \in \mathbb{R}$ tali per cui

$$\begin{cases} \nabla_x \mathcal{L}(x^*, \lambda, \lambda') = \nabla f(x^*) - \sum_{i=1}^p \lambda_i \nabla g_i(x^*) - \sum_{j=1}^q \lambda'_j \nabla h_j(x^*) = 0 \\ \lambda'_j \geq 0 \text{ per ogni } j = 1, \dots, q \\ \lambda'_j h_j(x^*) = 0 \text{ per ogni } j = 1, \dots, q. \end{cases}$$

A questo punto le condizioni KKT necessarie di ottimalità al second'ordine richiedono una migliore regolarità di ognuna delle funzioni in gioco e riguardano infatti la matrice hessiana della funzione lagrangiana - calcolata sempre in $x \in \mathbb{R}^n$: se f e le g_i, h_j son tutte di classe \mathcal{C}^2 su tutto \mathbb{R}^n , se $x^* \in \mathbb{R}^n$ è un punto ammissibile regolare e di minimo locale per f e se inoltre $\lambda_i, \lambda'_j \in \mathbb{R}$ sono come sopra dei $p+q$ moltiplicatori corrispondenti, allora la matrice

$$\nabla_x^2 \mathcal{L}(x^*, \lambda, \lambda') = \nabla^2 f(x^*) - \sum_{i=1}^p \lambda_i \nabla^2 g_i(x^*) - \sum_{j=1}^q \lambda'_j \nabla^2 h_j(x^*)$$

dev'essere semi-definita positiva su di un opportuno sottospazio V di \mathbb{R}^n che dipende da x^* stesso e da tutti i vincoli attivi g_i per $i = 1 \dots, p$ e h_{j_0} per $j_0 \in \mathcal{J}_0$, ma del quale non c'interessa meglio dettagliare l'aspetto.

Viceversa, finalmente, se esiste $x^* \in \mathbb{R}^n$ ammissibile e regolare e se esistono $p+q$ moltiplicatori non-negativi $\lambda_i, \lambda'_j \geq 0$ per i quali valgano le condizioni KKT necessarie di ottimalità al prim'ordine scritte sopra, più la proprietà

$$\forall j = 1, \dots, q, h_j(x^*) = 0 \Rightarrow \lambda'_j > 0,$$

e per i quali poi la matrice hessiana $\nabla_x^2 \mathcal{L}(x^*, \lambda, \lambda')$ risulti definita positiva su quel sottospazio V di \mathbb{R}^n legato a x^* e a tutti i vincoli attivi g_i per $i = 1 \dots, p$ e h_{j_0} per $j_0 \in \mathcal{J}_0$, allora x^* è un punto di minimo locale per la funzione obiettivo $f(x)$ - ed i λ_i, λ'_j possono esser considerati come i relativi moltiplicatori.

Di questa situazione, rimarchiamo in particolare che se tutte le funzioni f, g_i, h_j fossero pure convesse su \mathbb{R}^n allora x^* sarebbe automaticamente un punto di minimo globale per la f , cioè una soluzione del problema di partenza.

Tutti questi risultati teorici possono esser facilmente applicati al solito generico problema di QP su \mathbb{R}^n

$$\begin{cases} \min_x \frac{1}{2} x^T Q x + c^T x \\ A x = b \\ x \geq 0 : \end{cases}$$

dato infatti che per esso la funzione obiettivo è $f(x) = \frac{1}{2} x^T Q x + c^T x$ con la matrice Q semi-definita positiva, e dato poi che gli $m+n$ vincoli ai quali è soggetta sono tutti lineari, per ogni candidato punto ammissibile di minimo $x^* \in \mathbb{R}^n$ per f le condizioni KKT necessarie di ottimalità al prim'ordine collimano ad avere semplicemente

$$\begin{cases} Q x^* + c - A^T y - s = 0 \\ s \geq 0 \\ s_i x_i^* = 0 \text{ per ogni } i = 1, \dots, n \end{cases}$$

per degli opportuni moltiplicatori $y = (y_1, \dots, y_m)^T \in \mathbb{R}^m$ e $s = (s_1, \dots, s_n)^T \in \mathbb{R}^n$ - relativi rispettivamente alle m equazioni $Ax = b$ ed alle n disequazioni $x \geq 0$. Tali condizioni sono anche sufficienti qualora Q risulta in realtà definita positiva, o equivalentemente non-singolare, ed in tal caso x^* è l'unico punto ammissibile di minimo per f .

1.3 Programmazione Quadratica: il Metodo di Punto Interno di Karmarkar

Riepiloghiamo sommariamente quanto concerne la QP di ciò che abbiamo fin'ora visto: assegnati ancora $n, m \in \mathbb{N} \setminus \{0\}$, $Q^{n \times n}$ simmetrica e semi-definita positiva, $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$ e $b \in \mathbb{R}^m$, e ripreso il problema primario di QP su \mathbb{R}_x^n

$$\begin{cases} \min_x \frac{1}{2}x^T Qx + c^T x \\ Ax = b \\ x \geq 0, \end{cases}$$

sappiamo che il suo problema duale associato coincide col problema di QP su $\mathbb{R}_{(x,y,s)}^{n+m+n}$ dato in forma standard come

$$\begin{cases} \max_{x,y,s} b^T y - \frac{1}{2}x^T Qx \\ A^T y - Qx + s = c \\ x, s \geq 0, \end{cases}$$

ed in virtù delle varie condizioni KKT di ottimalità a riguardo abbiamo appreso che un punto $x^* \in \mathbb{R}^n$ ammissibile - $Ax^* = b$ e $x^* \geq 0$ - è di minimo globale per la funzione obiettivo $f(x) = \frac{1}{2}x^T Qx + c^T x$, $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$, solo se esistono degli speciali moltiplicatori $y = (y_1, \dots, y_m)^T \in \mathbb{R}^m$ e $s = (s_1, \dots, s_n)^T \in \mathbb{R}^n$ tali che

$$\begin{cases} Qx^* + c - A^T y - s = 0 \\ s \geq 0 \\ s_i x_i^* = 0 \text{ per ogni } i = 1, \dots, n, \end{cases}$$

e che invece, al contrario, se Q è definita positiva e se $x^*, s \in \mathbb{R}^n$ e $y \in \mathbb{R}^m$ soddisfano alle condizioni

$$\begin{cases} Ax^* = b, x^* \geq 0 \\ Qx^* + c - A^T y - s = 0 \\ s \geq 0 \\ s_i x_i^* = 0 \text{ per ogni } i = 1, \dots, n \end{cases}$$

allora viceversa x^* è l'unico punto ammissibile di minimo per f , ovvero è l'unica soluzione del problema primario. Supponiamo pertanto ed una volta per tutte che la matrice Q sia definita positiva, oltre che simmetrica.

Veniamo finalmente al metodo di punto interno di Karmarkar, di acronimo inglese *IPM*, per questo problema di QP: come già anticipato in precedenza, tale metodo poggia pesantemente le sue basi proprio sulle condizioni KKT appena richiamate, e procede modellando passo dopo passo quelle che sostanzialmente sarebbero le iterazioni del metodo degli iperpiani tangenti di Newton applicato allo stesso problema in un modo tale per cui risultino sempre rispettati tutti i diversi vincoli, soprattutto queglii descrivibili sotto forma di disuguaglianze.

Tutto comincia infatti da un'opportuna riscrittura compatta delle suddette condizioni KKT: convenendoci sulla notazione del tipo $X := \text{diag}[x], S := \text{diag}[s] \in \mathbb{R}^{n \times n}$, quali che siano $x = (x_1, \dots, x_n)^T, s = (s_1, \dots, s_n)^T \in \mathbb{R}^n$, ed indicando poi con $e = e(n) := (1, \dots, 1)^T \in \mathbb{R}^n$ il vettore di n componenti tutte uguali a 1, le condizioni KKT equivalgono al sistema quadratico, quindi non-lineare, nonché quadrato di esattamente $n + m + n = 2n + m$ equazioni in altrettante incognite, dato da

$$F(x, y, s) := \begin{bmatrix} A^T y - Qx + s - c \\ Ax - b \\ XS e \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} x \\ s \end{bmatrix} \geq 0,$$

dove in particolare resta definito il campo vettoriale liscio $F(x, y, s) : \mathbb{R}^{n+m+n} \rightarrow \mathbb{R}^{n+m+n}$.

Sopra, per ogni $N \in \mathbb{N} \setminus \{0\}$ ed ogni $v = (v_1, \dots, v_N)^T \in \mathbb{R}^N$, con $\text{diag}[v]$ stiamo intendendo quella matrice diagonale $N \times N$ avente elementi diagonali $\text{diag}[v]_{i,i} = v_i$ per ogni $i = 1, \dots, N$.

Ecco dunque perfettamente naturale l'idea di utilizzare in qualche modo il classico metodo di Newton per la ricerca di un $(x^*, y^*, s^*) \in \mathbb{R}^{n+m+n}$ con $F(x^*, y^*, s^*) = 0$, se certo non fosse per le vere difficoltà create dai $2n$ vincoli di non-negatività sulle variabili x e s .

L'approccio del metodo di punto interno segue la seguente strategia per trattare queste disequazioni col ≥ 0 :

- (I) scegliere un $(x^0, y^0, s^0) \in \mathbb{R}^{n+m+n}$ tale che $A^T y^0 - Qx^0 + s^0 = c$ e $Ax^0 = b$, ma non tale che necessariamente $X^0 S^0 e = 0$ - ove $X^0 = \text{diag}[x^0]$ e $S^0 = \text{diag}[s^0]$ -, e con inoltre $x^0 > 0$ e $s^0 > 0$, per cui in particolare (x^0, y^0, s^0) è un punto interno della regione ammissibile;
- (II) generare $(x^k, y^k, s^k) \in \mathbb{R}^{n+m+n}$ per $k = 1, 2, \dots$, ovvero $k \in \mathbb{N} \setminus \{0\}$, che soddisfino sempre $A^T y^k - Qx^k + s^k = c$ e $Ax^k = b$, e che si avvicinino progressivamente per $k \rightarrow \infty$ a soddisfare pure $X^k S^k e = 0$ - ove $X^k = \text{diag}[x^k]$ e $S^k = \text{diag}[s^k]$ -, avvicinandosi cioè dalla parte interna della regione ammissibile verso il suo bordo.

Presentiamo per l'esattezza il cosiddetto metodo di punto interno *primario-duale* il quale risolve al tempo stesso il problema primario ed appunto il suo duale. Sebbene occuparsi di due problemi invece che di uno soltanto potrebbe in un primo momento sembrare un'inutile perdita di tempo e di energie, tale scelta comporta al contrario una maggiore efficienza e robustezza del metodo rispetto ad altri più classici, e ciò dovrebbe risultare piuttosto intuitivo: la conoscenza aggiuntiva dei punti $(y^k, s^k) \in \mathbb{R}^{m+n}$, $k \in \mathbb{N}$, non potrà che permettere d'individuare con più precisione e più velocità i punti $x^k \in \mathbb{R}^n$, $k \in \mathbb{N}$, ad essi strettamente legati.

Come per tutti gli algoritmi iterativi di ottimizzazione, abbiamo bisogno in primo luogo di una buona misura di ottimalità - una misura che bene valuti e confronti la qualità delle diverse soluzioni alternative ed altresì la ricerca delle possibili direzioni di movimento - ed in secondo luogo di un criterio per generare una soluzione iterata migliore.

Cominciamo intanto considerando il sottoinsieme \mathcal{F} di \mathbb{R}^{n+m+n} definito ponendo

$$\mathcal{F} := \{(x, y, s) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n : Ax = b, A^T y - Qx + s = c, x \geq 0, s \geq 0\}$$

e considerando quindi la sua parte interna \mathcal{F}^0 - che supponiamo $\neq \emptyset$ - data da

$$\mathcal{F}^0 := \{(x, y, s) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n : Ax = b, A^T y - Qx + s = c, x > 0, s > 0\}.$$

Se immaginassimo per il momento una situazione di assoluta mancanza di ogni vincolo col ≥ 0 , se optassimo allora di risolvere il sistema $F(x, y, s) = 0$ per mezzo del metodo di Newton classico - in $(x, y, s) \in \mathbb{R}^{n+m+n}$ -, e se inoltre la stima corrente (x^k, y^k, s^k) della reale soluzione (x^*, y^*, s^*) fosse esattamente nell'insieme \mathcal{F}^0 per un certo $k \in \mathbb{N}$, allora il passo successivo secondo Newton richiederebbe di porre $(x^{k+1}, y^{k+1}, s^{k+1}) := (x^k, y^k, s^k) + (\Delta x^k, \Delta y^k, \Delta s^k)$ dove il vettore di correzione $(\Delta x^k, \Delta y^k, \Delta s^k) \in \mathbb{R}^{n+m+n}$ viene determinato risolvendo l'equazione matriciale

$$J_F(x^k, y^k, s^k) \begin{bmatrix} \Delta x^k \\ \Delta y^k \\ \Delta s^k \end{bmatrix} = -F(x^k, y^k, s^k)$$

o cioè, più esplicitamente,

$$\begin{bmatrix} -Q & A^T & I_n \\ A & 0 & 0 \\ S^k & 0 & X^k \end{bmatrix} \begin{bmatrix} \Delta x^k \\ \Delta y^k \\ \Delta s^k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -X^k S^k e \end{bmatrix},$$

in quanto dalla definizione di F e da quella di \mathcal{F}^0 abbiamo subito che

$$J_F(x, y, s) = \begin{bmatrix} -Q & A^T & I_n \\ A & 0 & 0 \\ S & 0 & X \end{bmatrix} \text{ per ogni } (x, y, s) \in \mathbb{R}^{n+m+n}, \text{ e che } F(x, y, s) = \begin{bmatrix} 0 \\ 0 \\ X S e \end{bmatrix} \text{ per ogni } (x, y, s) \in \mathcal{F}^0.$$

Tuttavia la problematica riguarda ovviamente il fatto che in generale $x^{k+1} \not\geq 0$ oppure $s^{k+1} \not\geq 0$, ovvero che in generale usciremmo dalla regione \mathcal{F} , e non solo: per tentare di risolvere questa problematica risulterebbe fin troppo ingenua pure un'idea tipo di accorciamento del passo, nel senso di cercare ad ogni iterazione un parametro $\alpha_k \in (0, 1]$ piccolo abbastanza affinché risulti

$$\begin{cases} x^k + \alpha_k \Delta x^k > 0 \\ s^k + \alpha_k \Delta s^k > 0, \end{cases}$$

prendendo quindi l' α_k più grande fra quelli che fanno ciò, e porre piuttosto

$$(x^{k+1}, y^{k+1}, s^{k+1}) := (x^k, y^k, s^k) + \alpha_k (\Delta x^k, \Delta y^k, \Delta s^k).$$

Siamo costretti ad apportare invece una modifica sostanziale che riguarda la ricerca stessa della direzione ottimale: iterazione dopo iterazione ci muoveremo dalla parte interna \mathcal{F}^0 di \mathcal{F} verso il suo bordo puntando sempre per l'esattezza verso uno speciale cammino regolare \mathcal{C} tutto contenuto in \mathcal{F}^0 , chiamato il *cammino centrale* - dall'inglese *central path* - il quale si rivela essere un protagonista del IPM sia per lo studio teorico ché per l'implementazione del metodo stesso, ed il quale per comune convenzione completa infatti il nome del IPM in *IPM path-following*.

Tale traiettoria \mathcal{C} viene parametrizzata mediante un certo scalare $\tau > 0$ idealmente piccolo - vicino allo zero - e tutti i punti $(x_\tau, y_\tau, s_\tau) \in \mathcal{F}^0$ del sostegno di \mathcal{C} , ovvero tali che $\mathcal{C} = \{(x_\tau, y_\tau, s_\tau) : \tau > 0\}$, sono chiamati *punti centrali* e sono tutte e sole le soluzioni del seguente sistema non-lineare $(2n + m) \times (2n + m)$:

$$F(x_\tau, y_\tau, s_\tau) = \begin{bmatrix} 0 \\ 0 \\ \tau e \end{bmatrix}, \quad \begin{bmatrix} x_\tau \\ s_\tau \end{bmatrix} > 0.$$

Dunque la sola differenza col sistema di partenza è il terzo blocco $n \times n$ di equazioni del sistema, ossia $X_\tau S_\tau e = \tau e$, che equivale a $(x_\tau)_i (s_\tau)_i = \tau$ per ogni $i = 1, \dots, n$, ed è il motivo naturale secondo cui dev'essere $\tau \simeq 0$. In effetti,

essendo $\mathcal{F}^0 \neq \emptyset$, è possibile dimostrare che per ogni fissato $\tau > 0$ questo sistema ammette una ed una sola soluzione (x_τ, y_τ, s_τ) , la quale poi converge alla soluzione (x^*, y^*, s^*) del problema primario nel limite $\tau \downarrow 0$.

Una generica interpretazione grafica della situazione in questione è davvero immediata, ed il tutto porta alla semplice idea base che sta dietro la strategia risolutiva generale del IPM path-following di generare iterativamente punti in \mathcal{F}^0 che bene approssimino i punti centrali per valori decrescenti del parametro $\tau > 0$.

Le ragioni secondo le quali decidiamo di non risolvere direttamente il sistema iniziale - col $\tau = 0$ - quanto piuttosto di risolvere tanti altri sistemi che in verità c'interessano meno, per di più approssimando le soluzioni di questi sistemi - ovvero i punti centrali - invece che calcolandole esplicitamente con esattezza, sono ovvie: da una parte, anzitutto, non esiste metodo generale teoricamente buono e computazionalmente economico per risolvere il sistema iniziale, non-lineare a causa dei vincoli di equazioni $x_i s_i = 0$, $i = 1, \dots, n$, ed inoltre l'applicazione standard del metodo di Newton presenta vari guai a causa dei vincoli di non-negatività $x, s \geq 0$, mentre invece è evidente che rispettare passo passo le condizioni strette

$$\begin{cases} x > 0 \\ s > 0 \\ x_i s_i = \tau > 0 \end{cases}$$

porta a generare punti in \mathcal{F}^0 sufficientemente lontani dal bordo della regione \mathcal{F} , ed in particolare sempre interni. È esattamente in questo senso che le iterate del IPM path-following puntano a \mathcal{C} , il quale fa in definitiva da guida verso la soluzione ottima. D'altra parte, infine, l'approssimazione dei punti centrali è dovuta naturalmente a ragioni di efficienza computazionale, giacché per ogni $\tau \gtrsim 0$ il sistema da risolvere resta comunque simile a quello iniziale, tanto più quanto più $\tau \downarrow 0$, per cui la sua risoluzione o anche solo un'approssimazione davvero accurata della relativa soluzione comporta pressoché le medesime difficoltà del problema iniziale.

Sopraggiunge così anche l'idea di partire con approssimazioni corrispondenti a τ relativamente grandi, o almeno abbastanza grandi affinché le problematiche sopra descritte rimangano trascurabili, per poi trovare opportunamente un'approssimazione di (x_τ, y_τ, s_τ) per ciascun $\tau \downarrow 0$.

Pertanto andiamo finalmente a distinguere le direzioni che appunto seguono il cammino centrale da quelle imposte dal metodo di Newton classico - le quali, ripetiamolo, portano in generale a risultati piuttosto poveri, specialmente nelle vicinanze del bordo, dove può presentarsi un comportamento di eccessiva deviazione rispetto alla regione ammissibile: ogni passo del metodo di Newton nel IPM è chiamato *passo puro* quando è un passo avente direzione esattamente verso il punto $(x^*, y^*, s^*) \in \mathcal{F}$ che verifica $F(x^*, y^*, s^*) = 0$, mentre invece la direzione di un qualsiasi passo del IPM che punti dall'iterata corrente trovata verso un pre-determinato punto centrale è chiamata *direzione centrata*. La possibile differenza grafica fra queste due tipologie di direzione dovrebbe risultare immediata.

Una direzione centrata è quindi ottenuta applicando per certi $\tau > 0$ un aggiornamento del metodo di Newton al sistema quadratico, quindi non-lineare, e quadrato di $2n + m$ equazioni in altrettante incognite, dato da

$$\widehat{F}_\tau(x, y, s) := \begin{bmatrix} A^T y - Qx + s - c \\ Ax - b \\ XSe - \tau e \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} :$$

osservato che per ogni $\tau > 0$ il campo vettoriale liscio $\widehat{F}_\tau(x, y, s)$ di \mathbb{R}^{n+m+n} possiede $J_{\widehat{F}_\tau} = J_F$, si tratta di risolvere

$$\begin{bmatrix} -Q & A^T & I_n \\ A & 0 & 0 \\ S^k & 0 & X^k \end{bmatrix} \begin{bmatrix} \Delta x_c^k \\ \Delta y_c^k \\ \Delta s_c^k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \tau e - X^k S^k e \end{bmatrix},$$

dove il pedice “c” del vettore incognito di correzione sta per “centrale”.

A questo punto è cruciale proprio la scelta di τ , ed è per questo che adesso introduciamo una misura di ottimalità che faccia a caso nostro: definiamo il *gap di dualità* - o la *media delle complementarità* - fra $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ e $s = (s_1, \dots, s_n)^T \in \mathbb{R}^n$ il numero

$$\mu = \mu(x, s) := \frac{1}{n} \sum_{i=1}^n x_i s_i = \frac{1}{n} x^T s = \frac{1}{n} \langle x; s \rangle_2 .$$

Così, in effetti, quando un dato $(x, y, s) \in \mathbb{R}^{n+m+n}$ soddisfa a tutte le condizioni $Ax = b$, $x \geq 0$, $A^T y - Qx + s = c$ e $s \geq 0$, allora (x, y, s) è la soluzione ottima se e sole se $\mu(x, s) = 0$, ed altrimenti comunque $\mu(x, s) > 0$ e quanto più questo è grande - lontano dallo zero - tanto più (x, y, s) è lontano dall'essere ottimale, e pertanto μ può davvero esser considerata una misura di ottimalità per i punti ammissibili. Questo non è tutto, perché ora quanto accade in particolare per i punti centrali c'interessa anche di più - e ci permette anche di spiegare meglio il fattore $\frac{1}{n}$ presente nell'espressione di μ : infatti, per ogni (x_τ, y_τ, s_τ) centrale, viene chiaramente $\mu(x_\tau, s_\tau) = \frac{1}{n} \sum_{i=1}^n \tau = \tau$.

Da tutto questo l'idea per procedere con l'algoritmo: per ogni $\tau > 0$, associare al punto centrale (x_τ, y_τ, s_τ) ogni altro punto ammissibile $(x, y, s) \in \mathcal{F}^0$ avente $\mu(x_\tau, s_\tau) = \tau$, come se in un certo senso fossero di uno stesso livello, e quindi per scegliere una direzione centrale a partire da un'iterata corrente (x^k, y^k, s^k) prendiamo un punto centrale (x_τ, y_τ, s_τ) avente appunto $\tau \leq \mu(x^k, s^k)$ - idealmente $\tau < \mu(x^k, s^k)$.

Possiamo dunque descrivere il IPM path-following mediante il risolvere iterativamente un sistema della forma

$$\begin{bmatrix} -Q & A^T & I_n \\ A & 0 & 0 \\ S^k & 0 & X^k \end{bmatrix} \begin{bmatrix} \Delta x_c^k \\ \Delta y_c^k \\ \Delta s_c^k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \sigma^k \mu^k e - X^k S^k e \end{bmatrix}$$

dove $\mu^k := \mu(x^k, s^k)$, e dove $\sigma^k \in [0, 1]$ è una scelta quantità che descrive il rapporto del gap di dualità fra il punto centrale bersagliato ed il punto corrente - e tale che infatti $\tau = \tau^k = \sigma^k \mu^k$. Quando $\sigma^k = 1$, allora $\mu^k = \tau$ e la direzione centrata è in realtà pura, mentre per $\sigma^k = 0$ il metodo è di Newton classico: nella pratica vengono quasi sempre scelti σ^k strettamente intermedi a tali due estremi, ossia scelti in $(0, 1)$.

Rimarchiamo che per $\sigma^k \simeq 1$ non stiamo migliorando sensibilmente il gap di dualità, ma pure che tuttavia facendo passi più piccoli dovrebbe accadere di meglio centrare il bersagliato punto centrale.

Siamo finalmente pronti per descrivere il generico IPM path-following, il quale si appoggia costantemente sul metodo di Newton ma optando passo passo per le direzioni centrate:

(1) Scegliere un punto iniziale $(x^0, y^0, s^0) \in \mathcal{F}^0$, e per ogni $k = 0, 1, 2, \dots$ ripetere i due passi seguenti.

(2) Scegliere un parametro $\sigma^k \in [0, 1]$ di accentramento, porre $\mu^k = \frac{(x^k)^T s^k}{n}$ e risolvere il sistema

$$\begin{bmatrix} -Q & A^T & I_n \\ A & 0 & 0 \\ S^k & 0 & X^k \end{bmatrix} \begin{bmatrix} \Delta x^k \\ \Delta y^k \\ \Delta s^k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \sigma^k \mu^k e - X^k S^k e \end{bmatrix}.$$

(3) Scegliere un parametro $\alpha^k \in (0, 1]$ di accorciamento tale che $x^k + \alpha^k \Delta x^k > 0$ e $s^k + \alpha^k \Delta s^k > 0$. Porre quindi $(x^{k+1}, y^{k+1}, s^{k+1}) = (x^k, y^k, s^k) + \alpha^k (\Delta x^k, \Delta y^k, \Delta s^k)$ e porre $k = k + 1$.

1.3.1 Le Varianti Short-Step e Long-Step del Metodo di Punto Interno

Alcune varianti del IPM path-following sorgono semplicemente dal come vengono scelti ad ogni iterazione i due parametri σ^k e α^k , fermo restando che tutte le possibili scelte di questi dovrebbero esser ragionevoli ed accurate a tal punto da garantire ad ogni passo la generazione di buone approssimazioni dei punti centrali.

Più precisamente vorremmo che ad ogni passo dell'algoritmo e per certi $\tau > 0$ fossero generati punti in un intorno di una parte di \mathcal{C} della forma $\{(x, y, s) \in \mathcal{F}^0 : \|(x, y, s) - (x_\tau, y_\tau, s_\tau)\|_2 \leq \varepsilon\}$, dove naturalmente $\varepsilon \gtrsim 0$. Peccato che in generale non si conosca mai esplicitamente alcun punto centrale (x_τ, y_τ, s_τ) , avendone piuttosto solo una descrizione implicita dal relativo sistema, e per ciò una rappresentazione di vicinanza come quella sopra risulta quasi del tutto insignificante nella pratica: occorrono di conseguenza rappresentazioni ben più utili computazionalmente le quali perlomeno implichino la vicinanza ai punti centrali. Due possibilità fra le più comuni sono le seguenti:

- al variare di $\vartheta \in (0, 1)$, $\mathcal{N}_2(\vartheta) := \{(x, y, s) \in \mathcal{F}^0 : \|XSe - \mu e\|_2 \leq \vartheta \mu, \mu := \frac{x^T s}{n}\}$;
- al variare di $\gamma \in (0, 1)$, $\mathcal{N}_{-\infty}(\gamma) := \{(x, y, s) \in \mathcal{F}^0 : x_i s_i \geq \gamma \mu, \mu := \frac{x^T s}{n}\}$.

A loro riguardo possiamo infatti osservare immediatamente che anzitutto

$$\begin{cases} \text{per ogni } 0 < \vartheta_1 \leq \vartheta_2 < 1, \mathcal{C} \subseteq \mathcal{N}_2(\vartheta_1) \subseteq \mathcal{N}_2(\vartheta_2) \text{ ed inoltre } \mathcal{N}_2(\vartheta) \downarrow \mathcal{C} \text{ nel limite } \vartheta \downarrow 0 \\ \text{per ogni } 0 < \gamma_1 \leq \gamma_2 < 1, \mathcal{C} \subseteq \mathcal{N}_{-\infty}(\gamma_2) \subseteq \mathcal{N}_{-\infty}(\gamma_1) \text{ ed inoltre } \mathcal{N}_{-\infty}(\gamma) \downarrow \mathcal{C} \text{ nel limite } \gamma \uparrow 1 \end{cases}$$

e che poi, essendo per ogni $(x, y, s) \in \mathbb{R}^{n+m+n}$ avente $x^T s > 0$ - com'è per tutti i punti $(x, y, s) \in \mathcal{F}^0$ - che vale $\|XSe - \mu e\|_2 \leq \vartheta \mu$ se e solo se vale $\left\| \frac{XS}{\mu} e - e \right\|_2 \leq \vartheta$, ossia più esplicitamente

$$\left\| \begin{bmatrix} \frac{x_1 s_1}{\mu} - 1 \\ \vdots \\ \frac{x_n s_n}{\mu} - 1 \end{bmatrix} \right\|_2^2 = \sum_{i=1}^n \left(\frac{x_i s_i}{\mu} - 1 \right)^2 \leq \vartheta^2,$$

risulta d'immediata verifica pure l'inclusione $\mathcal{N}_2(\vartheta) \subseteq \mathcal{N}_{-\infty}(\gamma)$ valida per ogni $\gamma \leq 1 - \vartheta$ - come quando viene preso $\vartheta \simeq 0$ -, ed in definitiva \mathcal{C} è più piccolo di \mathcal{N}_2 e, moralmente, \mathcal{N}_2 è a sua volta più piccolo di $\mathcal{N}_{-\infty}$.

Così un qualsiasi dato $(x, y, s) \in \mathbb{R}^{n+m+n}$ giace in $\mathcal{N}_2(\vartheta)$, dove $\vartheta \simeq 0$, solo se giace in \mathcal{F}^0 e se la somma dei quadrati delle relative deviazioni è sufficientemente piccola, ed in particolare $\mathcal{N}_2(\vartheta)$ sarà in generale molto meno esteso rispetto all'intera regione \mathcal{F}^0 , sempre per $\vartheta \simeq 0$, pur circondando tutto il cammino centrale \mathcal{C} .

Al contrario la sola richiesta per l'appartenenza del generico $(x, y, s) \in \mathcal{F}^0$ a $\mathcal{N}_{-\infty}(\gamma)$, dove $\gamma \simeq 0$, è che ciascuno degli n prodotti $x_i s_i$ non sia troppo piccolo rispetto alla loro media μ , e per ciò $\mathcal{N}_{-\infty}(\gamma)$ tenderà tipicamente a riempire una gran frazione della regione \mathcal{F}^0 , sempre per $\gamma \simeq 0$, compreso tutto \mathcal{C} .

Pertanto, tornando alla pratica del IPM path-following, se facessimo in modo da restringere tutte le iterate dell'algoritmo a generare soluzioni approssimate che siano punti di un certo intorno tipo \mathcal{N}_2 di \mathcal{C} , allora avremmo di

sicuro poco spazio di movimento e tutti i vari passi diverranno via via sempre più corti, mentre se ci accontentassimo di mantenerci in un intorno di \mathcal{C} tipo $\mathcal{N}_{-\infty}$ potremmo senz'altro contare su una buona libertà di spostamento.

Un IPM path-following è chiamato *IPM path-following short-step* nel caso che tutte le iterate da esso generate appartengano ad un intorno tipo \mathcal{N}_2 di \mathcal{C} , ed invece è chiamato più in generale *IPM path-following long-step* quando tutte le iterate da esso generate appartengono ad un intorno tipo $\mathcal{N}_{-\infty}$ di \mathcal{C} .

La marcata libertà di movimento consentita da un IPM path-following long-step viene certamente pagata nel contesto dei relativi ragionamenti teorici di tipo worst-case: in linea di principio solo le iterate all'interno di un intorno tipo \mathcal{N}_2 del cammino centrale \mathcal{C} - ed in particolare vicine a \mathcal{C} stesso - presenteranno le proprietà più desiderabili - proprio quelle di \mathcal{C} -, tra le quali una rapida discesa alla soluzione ottima del problema primario, quando invece a priori per un intorno tipo $\mathcal{N}_{-\infty}$ di \mathcal{C} , possibilmente molto vasto, esiste l'eventualità che le varie iterate restino fin troppo ravvicinate al bordo della regione \mathcal{F} , una condizione questa dalla quale potrebbe conseguire una convergenza sì inesorabile ma gravemente lenta.

Tuttavia quello che solitamente accade nella pratica è che gli IPMs path-following long-step non soltanto lavorano molto bene, ma addirittura funzionano apprezzabilmente meglio degli IPMs path-following short-step. L'idea che potrebbe sostenere a livello intuitivo una tale constatazione empirica è che vi è in ogni caso un legame piuttosto stretto fra un intorno tipo $\mathcal{N}_{-\infty}$ ed uno tipo \mathcal{N}_2 di \mathcal{C} , quindi in definitiva un legame piuttosto stretto fra $\mathcal{N}_{-\infty}$ e \mathcal{C} medesimo, e contemporaneamente in $\mathcal{N}_{-\infty}$ troviamo tanto spazio a sufficienza quanto occorre per muoversi in maniera più ottimale possibile verso la soluzione ottima.

In conclusione vogliamo riproporre l'algoritmo del IPM path-following ma nella sua variante long-step:

(1)' Fissare $\gamma \in (0, 1)$ e fissare $0 < \sigma_{min} < \sigma_{max} < 1$. Scegliere un punto iniziale $(x^0, y^0, s^0) \in \mathcal{N}_{-\infty}(\gamma)$, e per ogni $k = 0, 1, 2, \dots$ ripetere i due passi seguenti.

(2)' Scegliere un parametro $\sigma^k \in [\sigma_{min}, \sigma_{max}]$ di accentramento, porre $\mu^k = \frac{(x^k)^T s^k}{n}$ e risolvere il sistema

$$\begin{bmatrix} -Q & A^T & I_n \\ A & 0 & 0 \\ S^k & 0 & X^k \end{bmatrix} \begin{bmatrix} \Delta x^k \\ \Delta y^k \\ \Delta s^k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \sigma^k \mu^k e - X^k S^k e \end{bmatrix}.$$

(3)' Scegliere un parametro $\alpha^k \in (0, 1]$ di accorciamento tale che $(x^k, y^k, s^k) + \alpha^k (\Delta x^k, \Delta y^k, \Delta s^k) \in \mathcal{N}_{-\infty}(\gamma)$. Porre quindi $(x^{k+1}, y^{k+1}, s^{k+1}) = (x^k, y^k, s^k) + \alpha^k (\Delta x^k, \Delta y^k, \Delta s^k)$ e porre $k = k + 1$.

Osservazione. Un metodo di punto interno prescrive in ogni caso un'inizializzazione data dalla scelta di un qualsiasi $(x^0, y^0, s^0) \in \mathcal{F}^0$, un punto per così dire strettamente ammissibile, ma purtroppo non è sempre un compito così banale quello d'individuare uno, per cui siamo in generale costretti ad andare approssimando l'ottimalità ricercando contemporaneamente l'ammissibilità.

Con questo naturalmente intendiamo di cominciare da un arbitrario punto $(x^0, y^0, s^0) \in \mathbb{R}^{n+m+n}$ eventualmente non ammissibile ma avente perlomeno $x^0 > 0$ e $s^0 > 0$ - questo si generabile in modo davvero triviale - per procedere quindi con le solite iterate del metodo al quale apportiamo però una semplice modifica al sistema del passo (2)/(2)': per ogni dato $\tau > 0$, il generico passo del metodo di Newton applicato al sistema

$$\widehat{F}_\tau(x, y, s) := \begin{bmatrix} A^T y - Qx + s - c \\ Ax - b \\ XSe - \tau e \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad x \geq 0, \quad s \geq 0$$

corrisponde a risolvere da una certa iterata corrente $(x^k, y^k, s^k) \in \mathbb{R}^{n+m+n}$ - con $x^k > 0$ e $s^k > 0$ - il sistema

$$J_{\widehat{F}_\tau}(x^k, y^k, s^k) \begin{bmatrix} \Delta x^k \\ \Delta y^k \\ \Delta s^k \end{bmatrix} = -F(x^k, y^k, s^k),$$

al quale non richiediamo di presentare necessariamente le prime $n + m$ equazioni uguali a zero fin appunto dal principio dell'algoritmo, essendo possibilmente $(x^k, y^k, s^k) \notin \mathcal{F}^0$ per qualche $k \in \mathbb{N}$ iniziale, ed il quale diventa infatti

$$\begin{bmatrix} -Q & A^T & I_n \\ A & 0 & 0 \\ S^k & 0 & X^k \end{bmatrix} \begin{bmatrix} \Delta x^k \\ \Delta y^k \\ \Delta s^k \end{bmatrix} = \begin{bmatrix} c + Qx^k - A^T y^k - s^k \\ b - Ax^k \\ \tau e - X^k S^k e \end{bmatrix}.$$

2 Ottimizzazione di Portafoglio

2.1 Introduzione

Sappiamo che la finanza matematica va ormai evolvendosi in modo sempre più tecnico e sofisticato tanto nella teoria della ricerca quanto nella concretezza della pratica, e che le radici di un trend del genere potrebbero esser rintracciate in primo luogo nei metodi e nei modelli di strategia di portafoglio introdotti da H. Markowitz nei primi anni '50 e successivamente arricchiti da M. H. Miller e W. F. Sharpe - tutti vincitori del premio Nobel per l'economia

nel 1990, come già discusso a pagina 1 -, e quindi in secondo luogo nelle eleganti formule esplicite di pricing delle opzioni sviluppate da F. Black, M. Scholes e R. Merton negli anni '60 e '70 - le quali valsero a loro volta il premio Nobel per l'economia nel 1997 a Scholes e Merton (Black scomparve nell'agosto del 1995).

Il nostro lavoro verte senz'altro sul problema della strategia di portafoglio e conduce presto a capire che una scelta intelligente di una tale strategia debba essere anzitutto una scelta di diversificazione degli investimenti sui diversi possibili beni: infatti i veri protagonisti in gioco saranno i co-movimenti di tali beni, o cioè le correlazioni o semplicemente le covarianze relative ai profitti aleatori associati a ciascun bene - per una prefissata data futura.

Una diversificazione in senso proprio dovrebbe consistere nell'investire in beni i quali non necessariamente siano in quantità straordinariamente elevata, quanto invece che appartengano a dei settori piuttosto scorrelati fra loro.

In particolare si riveleranno necessariamente come protagonisti pure i rendimenti attesi di tali beni ed il rischio corrispondente ad ognuna di queste stime di guadagno.

Consideriamo per questo un certo investitore che vuole distribuire un dato ammontare unitario 1 di denaro contante investendolo in una certa quantità $n \in \mathbb{N} \setminus \{0, 1\}$ di diversi valori S_1, \dots, S_n , come potrebbero essere ad esempio stocks o bonds, e dai quali naturalmente avrà un ricavo aleatorio che spera esser a suo favore.

Rappresentiamo per ogni $i = 1, \dots, n$ un possibile fondo totale investito nel titolo S_i per mezzo di una variabile reale non-negativa $x_i \in \mathbb{R}_+$ e rappresentiamo quindi la corrispondente strategia di portafoglio come $x = (x_1, \dots, x_n)^T \in \mathbb{R}_+^n$. Allora in particolare il generico ricavo aleatorio risultante coincide col generico valore della v.a.r. $\sum_{i=1}^n S_i x_i$, ed inoltre $e^T x = \sum_{i=1}^n x_i = 1$ se ad $e = e(n) := (1, \dots, 1)^T \in \mathbb{R}^n$ lasciamo indicare di nuovo il vettore di n componenti tutte uguali a 1.

Per ciascuno di questi valori S_1, \dots, S_n poniamo che l'investitore si aspetti un ricavo di valore atteso e di deviazione standard conosciuti e dati rispettivamente da

$$\mu_i := \mathbf{E}[S_i] \in \mathbb{R} \quad \text{e} \quad \sigma_i := \mathbf{Var}[S_i]^{\frac{1}{2}} = (\mathbf{E}[S_i^2] - \mathbf{E}[S_i]^2)^{\frac{1}{2}} \geq 0, \quad i = 1, \dots, n,$$

quindi di varianza $\sigma_i^2 := (\sigma_i)^2 \geq 0$, o cioè in altri termini che si aspetti che all' i -esimo bene S_i resti associato a rischio di $\sigma_i \geq 0$ un rendimento atteso di μ_i . Indichiamo con $\mu := (\mu_1, \dots, \mu_n)^T \in \mathbb{R}^n$.

Supponiamo inoltre che per ogni coppia $i, j = 1, \dots, n$ sia noto il valore della covarianza fra i beni S_i e S_j

$$\sigma_{i,j} := \mathbf{Cov}[S_i, S_j] = \mathbf{E}[S_i S_j] - \mathbf{E}[S_i] \mathbf{E}[S_j]$$

e che sia quindi noto pure il valore del coefficiente di correlazione $\rho_{i,j}$ sempre fra S_i e S_j dato per definizione da

$$\rho_{i,j} := \frac{\sigma_{i,j}}{\sigma_i \sigma_j} = \frac{\mathbf{Cov}[S_i, S_j]}{\sigma_i \sigma_j}.$$

Indichiamo dunque con $\Sigma := (\sigma_{i,j})_{i,j=1,\dots,n} = (\rho_{i,j} \sigma_i \sigma_j)_{i,j=1,\dots,n} \in \mathbb{R}^{n^2}$, per cui certamente Σ è una matrice simmetrica e semi-definita positiva. In verità possiamo supporre che Σ sia non-singolare, ovvero definita positiva, in conseguenza del fatto che i titoli S_1, \dots, S_n possono esser assunti come non ridondanti fra loro.

Pertanto, confondendo la generica strategia di portafoglio $x = (x_1, \dots, x_n)^T \in \mathbb{R}_+^n$ sugli S_1, \dots, S_n col ricavo aleatorio risultante $\sum_{i=1}^n S_i x_i$, vediamo così che il portafoglio x ha valore atteso e varianza dati rispettivamente da

$$\mathbf{E}[x] = \sum_{i=1}^n \mu_i x_i = \mu^T x \quad \text{e} \quad \mathbf{Var}[x] = \mathbf{Cov}[x, x] = \sum_{i,j=1}^n \sigma_{i,j} x_i x_j = x^T \Sigma x.$$

In quest'ambito un dato portafoglio è chiamato *portafoglio efficiente* se presenta il massimo valore atteso fra tutti quei portafogli aventi la medesima varianza, o equivalentemente se presenta la minima varianza fra tutti i portafogli aventi perlomeno un certo valore atteso. I portafogli efficienti possono esser considerati come tutti e soli i portafogli interessanti, ed il loro insieme costituisce la cosiddetta *frontiera efficiente* dell'universo \mathbb{R}^n dei portafogli.

Il preannunciato modello di ottimizzazione media-varianza di Markowitz, o *modello MVO* di Markowitz, è un modello uno-periodale che si propone di risolvere il problema di strategia di portafoglio ottimo risolvendo proprio un generico problema di strategia di portafoglio efficiente - da cui il nome del modello stesso -, e sebbene se ne possa dare tre formulazioni differenti ma fra loro equivalenti resta comunque privilegiata quella sua versione come problema di QP su \mathbb{R}_x^n e descritta come segue: pre-assegnato uno scalare $R > 0$, risolvere

$$\begin{cases} \min_x x^T \Sigma x \\ e^T x = 1 \\ \mu^T x \geq R \\ x \geq 0. \end{cases}$$

Il modello MVO di Markowitz formalizza dunque il problema di portafoglio in un problema di QP che sappiamo ormai ben trattare, specie se appunto Σ fosse invertibile, grazie per esempio ad un algoritmo a tempo polinomiale tipo IPM path-following long-step con opportuna inizializzazione.

Per ciò uno dei pregi più notevoli di questo modello diviene evidentemente quello di una buona gestione del rischio - componente tanto inerente ad ogni attività finanziaria che si proponga guadagni netti piuttosto consistenti

quanto difficile da trattare anche col supporto di tecniche matematiche più avanzate - almeno quando una scelta misura di rischio $RM[\cdot]$ della generica strategia di portafoglio x risulta della forma $RM[x] \simeq x^T \Sigma x$.

Tuttavia nella pratica manifesta almeno due deprecabili difetti di natura intrinseca, specie su larga scala - per n elevato: il primo consiste in tendenziali sovrastime di quei titoli che sembrano presentare da una parte un buon rendimento atteso ad un rischio piuttosto basso, e d'altra parte una scarsa correlazione cogli altri titoli - penalizzando ogni rendimento che sia troppo diverso da quello atteso; infine il secondo consiste in una grave instabilità del problema stesso, nel senso che a delle variazioni sui dati di input μ_i , σ_i e $\sigma_{i,j}$, le quali parrebbero poter esser ritenute davvero piccole se non trascurabili, corrispondono invece delle variazioni solitamente fin troppo marcate su alcuni dei valori di output $\mathbf{E}[x]$, $\mathbf{Var}[x]$ e gli stessi x_i di composizione generale del portafoglio.

Ecco compresa a sufficienza la vera necessità di perfezionamenti più o meno importanti del modello MVO, pur mantenendosi nella classe dei modelli finanziari uno-periodali, dei quali forse il più riuscito è stato ideato da F. Black e R. Litterman all'interno della prestigiosa banca Goldman Sachs Group di New York nei primi anni '90, e prende infatti il nome di *modello Black-Litterman* o semplicemente *modello BL*: come capiremo a breve, il suo merito sostanziale è quello di dare un ruolo più rilevante sia all'equilibrio di mercato sia alle possibili previsioni che l'investitore può avere sull'andamento del mercato stesso, con tanto di grado di fiducia a riguardo.

Prima di passare ad approfondire la tecnica del modello MVO conviene pertanto rammentare alcuni dei concetti finanziari più fondamentali. Adesso invece concludiamo la sezione mostrando un possibile limite della MVO in qualità di statico modello uno-periodale, ovvero di modello non propriamente multi-periodale, all'interno del contesto della gestione delle attività - che danno profitto - assieme alle molteplici passività - che sono obblighi di pagamento - di dati beni, nonostante che da eventuali disallineamenti fra attività e passività nascano naturalmente dei rischi.

Immaginiamo per ciò che la variabile L_t rappresenti la passività - in inglese liability - dell'investitore o di una certa azienda al tempo t , al variare di $t = 1, \dots, T$ dove $T \in \mathbb{N} \setminus \{0\}$ è stato pre-fissato, ed immaginiamo per la precisione che ciascuna di queste L_t sia una v.a.r. di distribuzione nota.

Il tipico problema a riguardo consiste nel determinare su quali beni ed in quali quantità l'investitore o l'azienda dovrebbe investire in ogni periodo con lo scopo di massimizzare la ricchezza stimata alla fine del periodo T .

Assumendo che pure le classi $i = 1, \dots, n$ dei beni S_1, \dots, S_n fra le quali l'investitore o l'azienda possono scegliere abbiano tassi d'interesse aleatori $r_{i,t}$ di leggi note per ciascun periodo $t = 1, \dots, T$, l'idea naturale è quella di intraprendere decisioni in ogni periodo una volta osservato sia il profitto dai beni che l'ingenza delle passività del periodo immediatamente precedente, e pertanto il corrispondente problema d'ottimizzazione su \mathbb{R}^n potrebbe risultare non-deterministico e ricorsivo: un problema ad esempio della forma

$$\begin{cases} \max_x \mathbf{E}[\sum_{i=1}^n S_i x_{i,T}] \\ \sum_{i=1}^n (1 + r_{i,t}) S_i x_{i,t-1} - \sum_{i=1}^n S_i x_{i,t} = L_t \text{ per ogni } t = 1, \dots, T \\ x_{i,t} \geq 0 \text{ per ogni } i = 1, \dots, n \text{ e } t = 1, \dots, T \end{cases}$$

dove ovviamente intendiamo che per ogni $i = 1, \dots, n$ siano pre-fissati degli $x_{i,0} \geq 0$ - possibilmente anzi > 0 .

Osserviamo che effettivamente in ciascun periodo $t = 1, \dots, T$ il residuo lasciato in seguito a quando la passività L_t è stata coperta viene vincolato ad esser re-investito investendo $x_{i,t} \geq 0$ nella i -esima classe di bene: visto infatti che la quantità $\sum_{i=1}^n (1 + r_{i,t}) S_i x_{i,t-1}$ coincide col valore del portafoglio arrivati all'istante t mentre la quantità L_t coincide con quanto è d'obbligo pagare sempre al tempo t , la loro differenza corrisponderà a quanto possiamo re-investire sulla frazione scelta del mercato e la vincoliamo appunto ad essere uguale a $\sum_{i=1}^n S_i x_{i,t}$.

2.2 Preliminari: Alcuni Concetti Finanziari Fondamentali

Proponiamo un breve ripasso in una forma veramente schematica e con due scopi principali: il primo scopo è quello più serio e consiste nell'arrivare a ben ricordare concetto dopo concetto il *Capital Asset Pricing Model* di W. Sharpe, modello che c'interessa soprattutto perché in linea di principio potrebbe intersecare quello di MVO di Markowitz in via piuttosto importante, come avremo presto modo di vedere.

Il secondo scopo consiste invece di passare dal semplice meccanismo delle classiche *opzioni call* ed *opzioni put*, e quindi dai concetti di *pricing* - prezzaggio - e di *hedging* - copertura -, per presentare infine sotto forma di semplice esempio un altro grazioso problema di portafoglio: procediamo.

Rendimento privo di rischio. Un *rendimento privo di rischio*, o a *rischio zero*, è un rendimento del tutto sicuro, ovvero deterministico e nettamente > 0 . Un elementare esempio è dato dal rendimento dei Titoli di Stato a breve termine.

Rischio sistematico. Il *rischio sistematico* di una certa attività finanziaria è il fattore di rischio associato all'andamento del mercato nel suo complesso ed in generale non coincide esattamente col rischio complessivo dell'attività stessa, quanto piuttosto con quella componente del rischio totale che non può esser eliminata mediante strategie di diversificazione.

Rischio idiosincratico. Il *rischio idiosincratico* di una certa attività finanziaria rappresenta il rischio peculiare di tale attività ed è al contrario proprio quella componente del rischio complessivo che può esser ridotta per diversificazione.

Portafoglio di mercato. Il *portafoglio di mercato* è il portafoglio che comprende tutte le attività di mercato e dove i pesi di ciascuna attività, chiamati *pesi di capitalizzazione*, vengono rappresentati dalla capitalizzazione di mercato dell'attività stessa divisa per la capitalizzazione totale di tutte le attività rischiose.

Il portafoglio di mercato dovrebbe includere tutti i titoli azionari - le azioni - e tutti i titoli obbligazionari - i bonds - negoziati sia sui mercati organizzati che su quelli *over-the-counter*, o brevemente *OTC* - cioè quelli non quotati in borsa, dalla traduzione letterale "sul bancone" - ed altresì le attività non finanziarie - come i beni immobili, quelli di consumo durevole, ecc. - nonché il capitale umano, e pertanto l'esatta composizione del portafoglio di mercato non può certamente esser osservata.

Sappiamo bene che l'equilibrio di mercato richiede che la domanda delle attività rischiose sia uguale all'offerta delle attività rischiose: la domanda di attività rischiose viene rappresentata dal portafoglio possibilmente ottimo scelto dagli investitori, mentre l'offerta di attività rischiose viene rappresentata appunto dal portafoglio di mercato.

Premio al rischio. Un *premio al rischio* è un extra-rendimento rispetto al tasso di un'attività a rischio zero, il quale è richiesto da un investitore giustamente avverso al rischio per esser comunque incentivato ad acquistare un titolo rischioso, in virtù dell'ovvia idea di base che per acquistare titoli rischiosi bisogna venir incentivati da possibili guadagni interessanti. Dunque un premio al rischio può esser definito come la differenza fra il rendimento atteso di una data attività finanziaria ed il tasso d'interesse privo di rischio.

Naturalmente non tutto il rischio di un certo bene viene in questo senso ricompensato dal mercato, bensì soltanto quella parte di rischio che non può esser ridotta mediante scelte di diversificazione, ovvero il rischio sistematico associato.

Comunque in particolare il cosiddetto *premio al rischio di mercato* viene definito come la differenza fra il rendimento di mercato ed ancora il tasso d'interesse privo di rischio. Solitamente il premio al rischio di mercato può esser considerato come dato e varia tra il 5% ed il 7%.

Capital Asset Pricing Model. Il *Capital Asset Pricing Model*, o brevemente *CAPM*, è un modello teorico un-periodale proposto da W. Sharpe nei primi anni '60 per il calcolo del prezzo d'equilibrio di una certa attività finanziaria, secondo cui il rendimento atteso di tale attività è una funzione lineare del rendimento privo di rischio e del rischio sistematico all'attività stessa moltiplicato per il premio al rischio di mercato.

Secondo il CAPM, due attività che in situazione d'equilibrio risultino avere lo stesso rendimento atteso devono presentare pure lo stesso rischio sistematico nonostante che il rischio idiosincratice - e di conseguenza quello complessivo - delle due attività possa benissimo esser diverso.

In questo modello il rischio sistematico viene rappresentato dal cosiddetto *coefficiente beta* dell'attività ed è un numero ≥ 0 indicato appunto con β , mentre il rischio complessivo viene semplicemente misurato da una classica deviazione standard $\sigma \geq 0$.

Il CAPM richiede l'impiego di una proxy del portafoglio di mercato, ossia di un portafoglio non propriamente interessante di per sé se non per il fatto che risulti almeno strettamente correlato a quello di mercato. Una proxy utilizzata di frequente in questo senso è costituita dagli indici azionari - ossia da delle sintesi del valore di portafoglio dei titoli azionari che rappresentano - come ad esempio S&P/Mib per il mercato italiano o S&P500 per il mercato statunitense.

All'interno di questo modello il portafoglio di mercato coincide esattamente col portafoglio d'equilibrio del CAPM stesso, ed in particolare è composto dai beni in misura proporzionale alla loro effettiva capitalizzazione di mercato. La principale conclusione del CAPM risiede in effetti proprio nell'identificazione del portafoglio di mercato come portafoglio di tangenza tra domanda ed offerta in equilibrio.

In definitiva, finalmente, secondo le ipotesi del CAPM l'equilibrio tra rischio e rendimento di una data attività finanziaria i può esser espresso attraverso la seguente relazione lineare attesa fra il rendimento di tale titolo ed il rendimento del portafoglio di mercato:

$$\mathbf{E}[r_i] = \beta_{i,m}(\mathbf{E}[r_m] - r_f) + r_f ,$$

dove r_i rappresenta il rendimento lordo di tale i -esima attività, dove r_f indica il rendimento privo di rischio, dove r_m rappresenta il rendimento di mercato, e dove $\beta_{i,m}$ è il coefficiente beta sempre della i -esima attività ed è proporzionale per sua definizione alla covarianza fra il rendimento del titolo e l'andamento del mercato - motivo per il quale è davvero una misura di quanto il valore si muova in sintonia col mercato stesso -, avendo infatti espressione data precisamente da

$$\beta_{i,m} := \frac{\mathbf{Cov}[r_i, r_m]}{\mathbf{Var}[r_m]} .$$

In particolare il coefficiente beta del mercato assume il valore massimo di 1, e la quantità $\mathbf{E}[r_i] - r_f = \beta_{i,m}(\mathbf{E}[r_m] - r_f)$ coincide col premio al rischio di mercato e come tale rappresenta la misura in cui il rendimento atteso del portafoglio di mercato è superiore al rendimento a rischio zero.

Strategia di arbitraggio. Una *strategia di arbitraggio* è una strategia di commercio che presenti un rischio pari allo zero di perdite future pur contando su del denaro contante iniziale in quantità > 0 . Si tratta equivalentemente di una strategia che non richiede affatto denaro contante iniziale e ha rischio zero di perdita, ma ha pure probabilità > 0 di profitto futuro netto.

Per un generico modello di mercato finanziario viene tipicamente assunta come ipotesi fondamentale la cosiddetta *ipotesi NA* di non-arbitraggio, ovvero l'impossibilità d'intraprendere qualsiasi strategia d'arbitraggio, e sotto tale assunzione il generico problema di *pricing* di un qualsiasi strumento finanziario diventa il problema di determinare un *prezzo equo* di questo strumento, ossia appunto un prezzo di non-arbitraggio.

Opzione call ed opzione put. Una *opzione call europea* è uno strumento finanziario derivato, dunque un contratto, tale per cui ad una pre-fissata data di scadenza il titolare dell'opzione stessa ha il diritto, ma non l'obbligo, di acquistare un prescritto bene, il *sottostante*, per una sua prescritta quantità, il *prezzo strike*.

Una *opzione put europea* è un titolo completamente analogo ma con la sola differenza che vien assicurato il diritto di vendere piuttosto che quello di acquistare.

Una *opzione call americana* ed una *opzione put americana* sono esattamente come quelle rispettive europee, ma eventualmente possono esser esercitate in un qualsiasi momento anteriore rispetto alla data di scadenza.

Le opzioni call e le opzioni put, europee o americane che siano, fanno indubbiamente parte delle più note e comuni opzioni finanziarie, ma certo non sono le uniche.

Comunque, poiché evidentemente una data opzione ha sempre un profitto che dipende dal valore del titolo sottostante, il suo prezzo deve dipendere pure dal valore sia corrente che futuro del sottostante stesso. Per determinare un prezzo equo di un'opzione sotto l'ipotesi NA dev'esser risolto un problema di pricing.

Quando vi è un buon modello probabilistico per la complicata evoluzione aleatoria del valore del sottostante, il pricing delle opzioni può esser risolto con sofisticate tecniche matematiche d'integrazione stocastica.

In ogni caso un problema di pricing di una data opzione viene spesso risolto attraverso la seguente strategia di *hedging* sotto l'ipotesi NA: determinare un particolare portafoglio di opportuni valori corrispondenti a prezzi conosciuti e tali per cui, se correttamente rivalutati nel tempo, producano esattamente lo stesso profitto dell'opzione in questione. In questo modo infatti tale speciale portafoglio e l'opzione di partenza avranno per costruzione lo stesso profitto, e per ipotesi NA possiamo concludere che pure all'istante iniziale debbano avere lo stesso preciso valore - altrimenti appunto sarebbe possibile intraprendere un'ovvia strategia di arbitraggio.

Un portafoglio di questo tipo è un *portafoglio replicante* o *portafoglio di copertura* per l'opzione considerata.

Prevalentemente le opzioni vengono usate per i due scopi ben differenti di *speculazione* e di *copertura*.

Un fenomeno di speculazione può manifestarsi quando un certo investitore immagina una ben precisa evoluzione del prezzo del bene sottostante ad una data opzione e prende di conseguenza una posizione per così dire "spoglia" in opzioni su tale bene. A cominciare infatti dalla constatazione di quanto le opzioni siano spesso decisamente più economiche rispetto al relativo sottostante, una tale scommessa gli renderebbe guadagni netti ben superiori, nel caso l'evoluzione dei prezzi andasse come da egli prevista, rispetto ai guadagni che gli si sarebbero presentati a prendere analoghe posizioni direttamente sul sottostante. Dopo di che resta vero che pure le conseguenti perdite potrebbero risultare davvero ingenti qualora la valutazione sull'andamento futuro si rivelasse invece errata.

Invece un procedimento di copertura è quello di ridurre il rischio finanziario della complessiva posizione di un investitore componendo un portafoglio di beni dal quale ci si possa aspettare la compensazione delle eventuali perdite. Come esempio possiamo pensare ad un investitore che possiede un titolo di una certa azienda XYZ ma che si sente piuttosto preoccupato del fatto che il prezzo di questo bene possa decadere significativamente: allora egli potrebbe acquistare un'opzione put su un titolo del genere e proteggersi così contro tutti gli eventuali livelli di prezzo che scendono al di sotto di una certa soglia inferiore o uguale al prezzo strike dell'opzione stessa. Infatti l'opzione costerà piuttosto poco rispetto al relativo sottostante, per cui se il bene non decadesse al di sotto del prezzo strike allora verrebbe mantenuto, mentre in caso contrario verrebbe esercitato il diritto put vendendo il titolo al prezzo strike, dato appunto che avrebbe ormai valore inferiore.

Terminiamo questa parte con un facile esempio di pricing di un'opzione call europea.

Esempio 3. Immaginiamo che un titolo di un certo stock dell'azienda XYZ sia attualmente valutato al prezzo noto di 40\$ e che invece il suo prezzo ad un mese da oggi sia aleatorio. Supponiamo però che questo suo valore possa solo o raddoppiarsi o dimezzarsi, ed inoltre che entrambe queste eventualità possano presentarsi con la medesima probabilità.

Allora possiamo descrivere la situazione grazie ad un elementare modello finanziario binomiale uno-periodale: il bene $S_0 = 40\$$ può evolversi o in $S_1(u) = 80\$$ - up - o in $S_1(d) = 20\$$ - down.

Ebbene, oggi acquistiamo un'opzione call europea per poter comprare un titolo dello stock in questione al prezzo strike K di 50\$ ed alla scadenza di un mese da oggi, sperando naturalmente che il valore di tale titolo incrementi a 80\$. Quale dovrebbe esser considerato dunque un prezzo equo di quest'opzione?

Osserviamo subito che il profitto dell'opzione è chiaramente di $(S_1 - K)^+ := \max\{S_1 - K, 0\}$, cioè o di 30\$ o di 0\$ nel caso rispettivamente degli esiti u e d .

Assumiamo l'ipotesi NA e per semplicità pure di poter prendere in prestito o prestare denaro contante senza alcun tasso d'interesse fra oggi ed il prossimo mese, ed inoltre di poter comprare o vendere ogni quantità dello stock di XYZ senza ad esempio alcuna commissione: supponiamo brevemente che il mercato sia senza attriti. Supponiamo poi che entro il prossimo mese l'azienda XYZ non debba pagare alcun dividendo - ovvero quella parte di utile che viene distribuito da una società ai propri azionisti.

Risolviamo dunque questo problema di pricing passando per il seguente problema di hedging: costruire oggi un portafoglio sul titolo stock sottostante - comprato o venduto - e sul denaro contante - preso in prestito o prestato - tale che il profitto dato dal portafoglio stesso ad un mese da oggi uguali esattamente il profitto dato dall'opzione.

Indichiamo quindi con $\Delta \in \mathbb{R}$ la quantità di titoli di XYZ e con $B \in \mathbb{R}$ la quantità di denaro contante detenuti dal generico candidato portafoglio di copertura: l'obiettivo è quello di determinare i Δ e B giusti.

Per questo osserviamo intanto che oggi il valore di tale portafoglio è di $P_0 = 40\Delta + B$ dollari. Grazie poi alle nostre assunzioni vediamo invece che ad un mese da oggi il valore di questo portafoglio diviene o di $P_1(u) = 80\Delta + B$ dollari o di $P_1(d) = 20\Delta + B$ dollari nel caso rispettivamente degli esiti u e d .

Pertanto l'opzione call viene replicata scegliendo Δ e B che soddisfano il seguente elementare sistema lineare 2×2 :

$$\begin{cases} 80\Delta + B = 30 \\ 20\Delta + B = 0 \end{cases}$$

È immediato verificare che questo sistema ammette una ed una sola soluzione coppia che è $\Delta = \frac{1}{2}$ e $B = -10$, ed in conclusione il valore di $P_0 = (40\Delta + B)\$ = 10\$$ coincide col desiderato prezzo equo dell'opzione.

Osservazione. Risulta semplice ma istruttiva la generalizzazione della precedente situazione: assegnati $u, d \in (0, \infty)$ con $u > d$, dato $r \geq 0$ come tasso d'interesse da oggi ad un mese e posto quindi $R := 1 + r$, abbiamo che il bene $S_0 > 0$ sottostante ad una certa opzione call europea su un titolo dell'azienda XYZ può diventare o $S_1^u = S_0u$ oppure $S_1^d = S_0d$. Dati allora i due possibili profitti futuri C_1^u e C_1^d di tale opzione relativi agli esiti u e d rispettivamente, troviamo un giusto prezzo iniziale C_0 dell'opzione sempre sotto l'ipotesi NA e di non-attrito per il mercato, oltre che sotto l'assunzione di assenza dei dividendi per XYZ, mediante la solita strategia di portafoglio di copertura composto sul titolo sottostante e sul denaro contante.

Ebbene, un portafoglio di valore iniziale $\Delta S_0 + B$ dollari assume dopo un mese il valore o di $\Delta S_0u + BR$ dollari o di $\Delta S_0d + BR$ dollari a seconda degli esiti u e d rispettivamente, per cui ovviamente vogliamo Δ e B tali che

$$\begin{cases} \Delta S_0u + BR = C_1^u \\ \Delta S_0d + BR = C_1^d \end{cases} :$$

otteniamo subito che $\Delta = \frac{C_1^u - C_1^d}{S_0(u-d)}$ e quindi che $B = \frac{uC_1^d - dC_1^u}{R(u-d)}$, e pertanto il prezzo equo di oggi dev'essere

$$C_0 = \Delta S_0 + B = \frac{C_1^u - C_1^d}{u-d} + \frac{uC_1^d - dC_1^u}{R(u-d)} = \frac{1}{R} \left[\frac{R-d}{u-d} C_1^u + \frac{u-R}{u-d} C_1^d \right].$$

Supponendo ora che $d < R < u$, abbiamo dunque che i due scalari $p_u := \frac{R-d}{u-d}$ e $p_d := \frac{u-R}{u-d}$ sono tali che $C_0R = p_u C_1^u + p_d C_1^d$, ed inoltre $p_u, p_d > 0$ e $p_u + p_d = 1$, ossia determinano una probabilità sugli esiti u e d del modello binomiale: questa è chiamata probabilità martingala equivalente, o probabilità risk-neutral - neutrale al rischio -, ed è indipendente dalle singole probabilità iniziali dei due stati.

Dunque sotto la probabilità risk-neutral il prezzo di non-arbitraggio dell'opzione considerata coincide esattamente col suo valore atteso futuro scontato al tasso d'interesse r , il quale è ovviamente privo di rischio: in altri termini tale attività finanziaria presenta sotto questa probabilità un tasso di rendimento atteso che è appunto a rischio zero, a prescindere invece dalla sua rischiosità iniziale.

2.3 Ottimizzazione di Portafoglio: il Modello MVO di Markowitz

Torniamo finalmente ad un investitore intenzionato a distribuire un ammontare unitario 1 di denaro contante su di una quantità $n \in \mathbb{N} \setminus \{0, 1\}$ di diversi valori S_1, \dots, S_n ciascuno dei quali presenti un rendimento atteso ed una deviazione standard conosciuti e dati rispettivamente da

$$\mu_i := \mathbf{E}[S_i] \in \mathbb{R} \quad \text{e} \quad \sigma_i := \mathbf{Var}[S_i]^{\frac{1}{2}} \geq 0, \quad i = 1, \dots, n,$$

per i quali indichiamo in particolare $\mu = (\mu_1, \dots, \mu_n)^T \in \mathbb{R}^n$ e $\sigma_i^2 := (\sigma_i)^2 \geq 0$ per ciascun $i = 1, \dots, n$.

Supponiamo inoltre che per ogni coppia $i, j = 1, \dots, n$ sia noto pure il valore della covarianza fra i beni S_i e S_j

$$\sigma_{i,j} := \mathbf{Cov}[S_i, S_j]$$

e che sia quindi noto pure il valore del coefficiente di correlazione sempre fra S_i e S_j dato invece da

$$\rho_{i,j} := \frac{\sigma_{i,j}}{\sigma_i \sigma_j},$$

per i quali indichiamo dunque $\Sigma := (\sigma_{i,j})_{i,j=1,\dots,n} = (\rho_{i,j} \sigma_i \sigma_j)_{i,j=1,\dots,n} \in \mathbb{R}^{n \times n}$ e per i quali conveniamo di supporre adesso ed una volta per tutte che tale Σ sia definita positiva.

Per ogni $i = 1, \dots, n$ rappresentiamo allora il fondo totale investito nel titolo S_i nuovamente mediante la variabile $x_i \in \mathbb{R}_+$, riferendoci così a $x = (x_1, \dots, x_n)^T \in \mathbb{R}_+^n$ come alla corrispondente strategia di portafoglio, e confondiamo tranquillamente x medesimo col ricavo aleatorio risultante $\sum_{i=1}^n S_i x_i$: pertanto abbiamo che in definitiva il portafoglio x ha valore atteso e varianza dati rispettivamente da

$$\mathbf{E}[x] = \sum_{i=1}^n \mu_i x_i = \mu^T x \quad \text{e} \quad \mathbf{Var}[x] = \sum_{i,j=1}^n \sigma_{i,j} x_i x_j = x^T \Sigma x .$$

Abbiamo già imparato che in questo contesto un portafoglio è considerato efficiente se presenta il massimo valore atteso fra tutti quei portafogli aventi la medesima varianza, o equivalentemente se presenta la minima varianza fra tutti i portafogli aventi perlomeno un certo valore atteso, e che la classe di tutti portafogli efficienti forma la frontiera efficiente dell'universo \mathbb{R}^n dei portafogli.

Naturalmente non tutti i portafogli risultano efficienti all'interno di un dato modello MVO, anzi: tra quelli che al contrario potremmo chiamare *portafogli inefficienti* troviamo generalmente già tutti i portafogli troppo poco diversificati, come ad esempio tutti quelli basati su di un unico singolo titolo.

Assumiamo a proposito che quella parte del totale dei portafogli che consideriamo ammissibili rispetto al problema di MVO di Markowitz sia descrivibile come un sottoinsieme non vuoto di \mathbb{R}^n e poliedrale del tipo

$$\chi := \{x \in \mathbb{R}^n : Ax = b, Cx \geq d\} ,$$

dove naturalmente sono da intendersi come dati $m, r \in \mathbb{N} \setminus \{0\}$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $C \in \mathbb{R}^{r \times n}$ e $d \in \mathbb{R}^r$.

In particolare un qualsivoglia vincolo lineare sul generico portafoglio x può venir sottinteso come già calcolato nel poliedro χ , come per esempio $e^T x = \sum_{i=1}^n x_i = 1$ - ove ancora $e = e(n) := (1, \dots, 1)^T \in \mathbb{R}^n$.

A questo punto, dato che per ipotesi la matrice Σ è definita positiva, la funzione varianza $\mathbb{R}^n \rightarrow \mathbb{R}_+$ data da $x \mapsto \mathbf{Var}[x] = x^T \Sigma x$ ristretta appunto agli $x \in \chi$ è sicuramente una funzione strettamente/fortemente convessa su di un sottoinsieme convesso, compatto e non vuoto di \mathbb{R}^n , e come tale ammette uno ed un solo punto di minimo, ossia uno ed un solo portafoglio ammissibile $x_{min} \in \chi$ tale per cui valga $\mathbf{Var}[x_{min}] \leq \mathbf{Var}[x]$ per ogni altro $x \in \chi$.

Poniamo quindi in sua corrispondenza $R_{min} := \mu^T x_{min} = \mathbf{E}[x_{min}] > 0$, notando in particolare che x_{min} è un portafoglio efficiente in modo ovvio, ed indichiamo invece con $R_{max} \in (0, \infty)$ il massimo valore atteso fra quelli dei portafogli ammissibili - teniamo sempre presente che in ogni caso l'insieme χ è un sottoinsieme compatto di \mathbb{R}^n .

Quanto più n è alto tanto più la varianza del generico portafoglio efficiente dipenderà dalle covarianze fra i titoli piuttosto che dalle loro singole varianze, ed infatti la matrice di covarianza potrebbe ovviamente esser considerata come indice del rischio sistematico dell'attività finanziaria globale risultante.

Discutiamo finalmente le tre formulazioni differenti ma fra loro equivalenti del modello MVO di Markowitz.

- (a) La prima formulazione, già anticipata poc'anzi all'interno dell'introduzione, è quella che riteniamo la migliore dal punto di vista computazionale ed è infatti quella che scegliamo adesso e per tutto il resto della sezione: si tratta del problema di QP su \mathbb{R}_+^n dato da

$$\begin{cases} \min_x x^T \Sigma x \\ \mu^T x \geq R \\ Ax = b \\ Cx \geq d \end{cases}$$

dove $R > 0$ è da considerarsi assegnato in modo tale per cui $R_{min} \leq R \leq R_{max}$.

Consiste proprio nel risolvere questo problema al variare di ogni R lungo tutto l'intervallo chiuso $[R_{min}, R_{max}]$ di $(0, \infty)$ che per definizione otteniamo tutti e soli i portafogli efficienti. Più precisamente infatti, se consideriamo la funzione $\sigma(R) : [R_{min}, R_{max}] \rightarrow \mathbb{R}$ definita ponendo $\sigma(R) := \mathbf{Var}[x_R]^{\frac{1}{2}} = (x_R^T \Sigma x_R)^{\frac{1}{2}}$, dove $x_R \in \chi$ è appunto la soluzione del problema sopra, allora la frontiera efficiente di χ è il grafico di σ in $[R_{min}, R_{max}] \times \mathbb{R}_+$:

$$E := \{(R, \sigma(R)) = (R, \mathbf{Var}[x_R]^{\frac{1}{2}}) : R \in [R_{min}, R_{max}]\} .$$

Inoltre per questo problema di QP le condizioni di ottimalità KKT al prim'ordine risultano sia necessarie che sufficienti e sono date come segue: per ogni fissato $R \in [R_{min}, R_{max}]$, un dato punto $x_R \in \mathbb{R}^n$ è la soluzione ottima del problema se e solo se esistono $1 + m + r$ moltiplicatori $\lambda_R \in \mathbb{R}$ - relativo alla disequazione $\mu^T x \geq R$ -, $\gamma_E \in \mathbb{R}^m$ - relativi alle m equazioni $Ax = b$ - e $\gamma_I \in \mathbb{R}^r$ - relativi alle r disequazioni $Cx = d$ - tali che

$$\begin{cases} \mu^T x_R \geq R, Ax_R = b, Cx_R \geq d \\ \Sigma x_R - \lambda_R \mu - A^T \gamma_E - C^T \gamma_I = 0 \\ \lambda_R \geq 0 \text{ e } \lambda_R (\mu^T x_R - R) = 0 \\ \gamma_I \geq 0 \text{ e } \gamma_I^T (Cx_R - d) = 0 . \end{cases}$$

- (b) Nella seconda formulazione vengono invece invertiti i ruoli del valore atteso e della varianza ottenendo di conseguenza un problema di massimizzazione su \mathbb{R}^n di una funzione addirittura lineare soggetta però anche ad un vincolo quadratico di disuguaglianza, e per ciò un problema che di certo non può essere di QP: pre-assegnato uno scalare $\sigma^2 > 0$, risolvere

$$\begin{cases} \max_x \mu^T x \\ x^T \Sigma x \leq \sigma^2 \\ Ax = b \\ Cx \geq d. \end{cases}$$

- (c) Infine la terza formulazione propone un naturale incrocio delle due precedenti mediante un problema di *programmazione conica* su \mathbb{R}^n : fissata una costante $\delta > 0$ di *avversione al rischio* - una misura cioè di quanto intendiamo preferire il valore atteso di una data quantità aleatoria rispetto alla quantità stessa -, risolvere

$$\begin{cases} \max_x \mu^T x - \frac{\delta}{2} x^T \Sigma x \\ Ax = b \\ Cx \geq d. \end{cases}$$

Introduciamo a questo punto un elementare esempio che c'accompagnerà da ora e via via fin quasi alla fine per supportarci nella comprensione dapprima del modello MVO di Markowitz nella sua prima formulazione (a) così com'è e quindi, a seguire, di quelle sue diverse varianti che sceglieremo a breve di presentare.

In effetti questo esempio attraversa in modo del tutto analogo buona parte dell'ottavo capitolo di (2) dal quale viene tratto ed è proprio alla consultazione di esso che invitiamo almeno per quanto concerne la risoluzione esplicita finale - mediante un qualche opportuno software - dei vari problemi di QP su \mathbb{R}^n che vi vengono ottenuti.

Procederemo invece in modo del tutto autonomo prima nell'estrapolazione a partire dalle serie storiche consegnate come dati di tutte quelle informazioni matematiche utili in modo fondamentale al modello, ricavandoci così da soli gli stessi identici numeri o quasi che presentano gli autori, ed infine nel tracciare i disegni grafici più significativi.

Faremo tutto questo servendoci del noto software statistico **R** su sistema operativo Mac OS X.

Esempio 4. *Il problema è quello della costruzione di un portafoglio efficiente basato sulle tre seguenti tipologie di valori: uno stock, un bond e del denaro contante di mercato monetario - l'insieme delle negoziazioni aventi ad oggetto prestiti monetari con durata tipicamente inferiore ai dodici mesi, una peculiarità questa che coincide con la differenza principale rispetto al mercato dei capitali. Consideriamo per ciò i seguenti dati storici per queste tre classi di beni, in numero di 44 per tipologia:*

Anno	Stocks	Bonds	MM
1960	20.2553	262.935	100.00
1961	25.6860	268.730	102.33
1962	23.4297	284.090	105.33
1963	28.7463	289.162	108.89
1964	33.4484	299.894	113.08
1965	37.5813	302.695	117.97
1966	33.7839	318.197	124.34
1967	41.8725	309.103	129.94
1968	46.4795	316.051	137.77
1969	42.5448	298.249	150.12
1970	44.2212	354.671	157.48
1971	50.5451	394.532	164.00
1972	60.1461	403.942	172.74
1973	51.3114	417.252	189.93
1974	37.7306	433.927	206.13
1975	51.7772	457.885	216.85
1976	64.1659	529.141	226.93
1977	59.5739	531.144	241.82
1978	63.4884	524.435	266.07
1979	75.3032	531.040	302.74
1980	99.7795	517.860	359.96
1981	94.8671	538.769	404.48
1982	115.308	777.332	440.68
1983	141.316	787.357	482.42
1984	150.181	907.712	522.84
1985	197.829	1200.63	566.08
1986	234.755	1469.45	605.20
1987	247.080	1424.91	646.17
1988	288.116	1522.40	702.77
1989	379.409	1804.63	762.16
1990	367.636	1944.25	817.87
1991	479.633	2320.64	854.10
1992	516.178	2490.97	879.04
1993	568.202	2816.40	905.06
1994	575.705	2610.12	954.39
1995	792.042	3287.27	1007.84
1996	973.897	3291.58	1061.15
1997	1298.82	3687.33	1119.51
1998	1670.01	4220.24	1171.91
1999	2021.40	3903.32	1234.02
2000	1837.36	4575.33	1313.00
2001	1618.98	4827.26	1336.89
2002	1261.18	5558.40	1353.47
2003	1622.94	5588.19	1366.73

Indicando quindi con $T := 43$ e con $t = 0, 1, \dots, T$ gli anni 1960, 1961, ..., 2003 rispettivamente, rappresentiamo i numeri della tabella come $I_{i,t}$ al variare di $i = 1, 2, 3$ e di $t = 0, 1, \dots, T$, per cui ciascun $I_{i,t}$ corrisponde al profitto totale del bene i al tempo t e, come tale, ci risulta un dato per così dire un po' troppo grezzo: convertiamo infatti ogni $I_{i,t}$ con $t \geq 1$ nel rispettivo tasso di rendimento $r_{i,t}$ mediante la formula fondamentale

$$r_{i,t} = \frac{I_{i,t} - I_{i,t-1}}{I_{i,t-1}},$$

ottenuta semplicemente ricordando che il generico $r_{i,t}$ è univocamente determinato dal fatto che $I_{i,t} = (1 + r_{i,t})I_{i,t-1}$.

Riportiamo qua sotto direttamente le tre serie storiche così ottenute, e già pronte quindi per esser agilmente caricate su **R** - a partire ad esempio dall'usuale combinazione di tasti `cmd+c` -, anziché la tabella corrispondente:

Stocks: 26.81 -8.78 22.69 16.36 12.36 -10.10 23.94 11.00 -8.47 3.94 14.30 18.99 -14.69 -26.47 37.23 23.93 -7.16 6.57 18.61
32.50 -4.92 21.55 22.56 6.27 31.17 18.67 5.25 16.61 31.69 -3.10 30.46 7.62 10.08 1.32 37.58 22.96 33.36 28.58 21.04
-9.10 -11.89 -22.10 28.68 ;

Bonds: 2.20 5.72 1.79 3.71 0.93 5.12 -2.86 2.25 -5.63 18.92 11.24 2.39 3.29 4.00 5.52 15.56 0.38 -1.26 -1.26 -2.48 4.04
44.28 1.29 15.29 32.27 22.39 -3.03 6.84 18.54 7.74 19.36 7.34 13.06 -7.32 25.94 0.13 12.02 14.45 -7.51 17.22 5.51
15.15 0.54 ;

MM: 2.33 2.93 3.38 3.85 4.32 5.40 4.51 6.02 8.97 4.90 4.14 5.33 9.95 8.53 5.20 4.65 6.56 10.03 13.78 18.90 12.37 8.95
9.47 8.38 8.27 6.91 6.77 8.76 8.45 7.31 4.43 2.92 2.96 5.45 5.60 5.29 5.50 4.68 5.30 6.40 1.82 1.24 0.98 .

Riserviamo inoltre per ogni $i = 1, 2, 3$ il simbolo R_i per i tassi aleatori del profitto dell' i -esimo bene.

Possiamo finalmente cominciare coi calcoli statistici in **R** a cominciare dalla media aritmetica dei tassi di profitto per ogni bene $i = 1, 2, 3$, ovvero da

$$\bar{r}_i := \frac{1}{T} \sum_{t=1}^T r_{i,t}, \quad i = 1, 2, 3,$$

evidenziando ordinatamente prima la serie storica degli stocks, poi la serie dei bonds ed infine quella del MM, copiando ogni volta ciascuna di esse mediante un semplice `cmd+c` e procedendo quindi per ognuna coi primi semplici comandi:

```
> Stocks <- scan(pipe("pbpaste"))
```

```
Read 43 items
```

```
> mean(Stocks)
```

```
[1] 12.04419
```

```
> Bonds <- scan(pipe("pbpaste"))
```

```
Read 43 items
```

```
> mean(Bonds)
```

```
[1] 7.792326
```

```
> MM <- scan(pipe("pbpaste"))
```

```
Read 43 items
```

```
> mean(MM)
```

```
[1] 6.323023
```

Riassumiamo in percentuale quanto appena ottenuto, sostituendo insomma \bar{r}_i con $\frac{\bar{r}_i}{100}$ per ogni $i = 1, 2, 3$:

	Stocks	Bonds	MM
Media Aritmetica	12.04%	7.79%	6.32%

Tuttavia i tassi di rendimento sono moltiplicativi nel tempo e per ciò dobbiamo preferire la media geometrica rispetto a quella aritmetica per la composizione del vettore μ del modello MVO. Questa media potrà quindi esser intesa come la costante di tasso annuale di rendimento per il calcolo del rendimento totale $I_{i,T}$ a partire da $I_{i,0}$, per ogni $i = 1, 2, 3$, applicandola via via dall'anno $t = 0$ all'anno $t = T - 1$. La formula che adoperiamo per questa media geometrica è

$$\mu_i := \left[\prod_{t=1}^T \left(1 + \frac{r_{i,t}}{100} \right) \right]^{\frac{1}{T}} - 1, \quad i = 1, 2, 3,$$

e quanto ricaviamo dunque da ciascuna delle tre serie storiche proseguendo in **R** dai precedenti comandi è qui riportato:

```
> prod(1+Stocks/100)^(1/43)-1
```

```
[1] 0.1072126
```

```
> prod(1+Bonds/100)^(1/43)-1
```

```
[1] 0.07304677
```

```
> prod(1+MM/100)^(1/43)-1
```

```
[1] 0.06270127
```

Sintetizziamo quanto ottenuto:

	Stocks	Bonds	MM
Media Geometrica	10.72%	7.3%	6.27%

A questo punto tutti gli \bar{r}_i , $i = 1, 2, 3$, occorrono invece nel calcolo della matrice $\Sigma = (\sigma_{i,j})_{i,j=1,2,3}$ delle covarianze fra stocks, bonds e MM, matrice 3×3 simmetrica e semi-definita positiva data componente per componente da

$$\sigma_{i,j} := \text{Cov}[R_i, R_j] = \frac{1}{T} \sum_{t=1}^T (r_{i,t} - \bar{r}_i)(r_{j,t} - \bar{r}_j), \quad i, j = 1, 2, 3 :$$

```
> A = matrix(nrow=43,ncol=3)
> A[,1] = Stocks
> A[,2] = Bonds
> A[,3] = MM
> cov(A/100)
      [,1]      [,2]      [,3]
[1,] 0.0283919768 0.0038850792 0.0002092916
[2,] 0.0038850792 0.0114793828 -0.0002448836
[3,] 0.0002092916 -0.0002448836 0.0011814812
> det(cov(A/100))
[1] 3.646344e-07
> det(cov(A))
[1] 364634.4
```

Pertanto la matrice Σ delle covarianze è in realtà definita positiva, e per chiarezza la riscriviamo di seguito:

	Stocks	Bonds	MM
Stocks	0.02839	0.00388	0.00021
Bonds	0.00388	0.01148	-0.00024
MM	0.00021	-0.00024	0.00118

Da essa possiamo da una parte dedurre immediatamente tutte le tre volatilità, ossia le deviazioni standard, date da

$$\sigma_i := \sqrt{\sigma_{i,i}} = \sqrt{\text{Cov}[R_i, R_i]}, \quad i = 1, 2, 3 :$$

```
> sd(Stocks/100)
[1] 0.1684992
> sd(Bonds/100)
[1] 0.1071419
> sd(MM/100)
[1] 0.03437268
```

Appuntiamocene tutte:

	Stocks	Bonds	MM
Volatilità	16.85%	10.71%	3.44%

Sempre da Σ deduciamo d'altra parte pure la matrice delle correlazioni calcolate per definizione come

$$\rho_{i,j} := \frac{\sigma_{i,j}}{\sigma_i \sigma_j} = \frac{\text{Cov}[R_i, R_j]}{\sigma_i \sigma_j}, \quad i, j = 1, 2, 3 :$$

```
> cor(A)
      [,1]      [,2]      [,3]
[1,] 1.00000000 0.21520026 0.03613603
[2,] 0.21520026 1.00000000 -0.06649469
[3,] 0.03613603 -0.06649469 1.00000000
```

Mettiamocela da parte:

	Stocks	Bonds	MM
Stocks	1	0.2152	0.0361
Bonds	0.2152	1	-0.0665
MM	0.0361	-0.0665	1

In conclusione, finalmente, il problema in questione di strategia di portafoglio ottimo può esser tradotto come problema di portafoglio efficiente su $\mathbb{R}^n = \mathbb{R}^3$ per mezzo del seguente modello MVO di Markowitz:

$$\begin{cases} \min_x x^T \Sigma x \\ \mu^T x \geq R \\ e^T x = 1 \\ x \geq 0 \end{cases}$$

dove $x := (x_S, x_B, x_M)^T \in \mathbb{R}^3$, $\Sigma = (\sigma_{i,j})_{i,j=1,2,3} \in \mathbb{R}^{3 \times 3}$, $\mu := (\mu_1, \mu_2, \mu_3)^T \in \mathbb{R}^3$ ed $e := (1, 1, 1)^T \in \mathbb{R}^3$, o esplicitamente

$$\begin{cases} \min_{(x_S, x_B, x_M)} 0.02839x_S^2 + 0.01148x_B^2 + 0.00118x_M^2 + \\ \quad + 2 \times 0.00388x_Sx_B + 2 \times 0.00021x_Sx_M - 2 \times 0.00024x_Bx_M \\ 0.1072x_S + 0.073x_B + 0.0627x_M \geq R \\ x_S + x_B + x_M = 1 \\ x_S \geq 0, x_B \geq 0, x_M \geq 0, \end{cases}$$

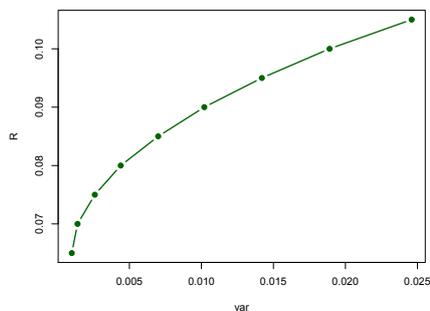
al variare ad esempio di $R > 0$ tra 6.27% \rightsquigarrow 6.5% e 10.72% \rightsquigarrow 10.5% con incrementi dello 0.5%.

Gli autori trovano la soluzione di questo problema di QP grazie all'ausilio di un software ed arrivano ai seguenti risultati:

R	Varianza ottima	Stocks	Bonds	MM
0.065	0.0010	0.03	0.10	0.87
0.070	0.0014	0.13	0.12	0.75
0.075	0.0026	0.24	0.14	0.62
0.080	0.0044	0.35	0.16	0.49
0.085	0.0070	0.45	0.18	0.37
0.090	0.0102	0.56	0.20	0.24
0.095	0.0142	0.67	0.22	0.11
0.100	0.0189	0.78	0.22	0
0.105	0.0246	0.93	0.07	0

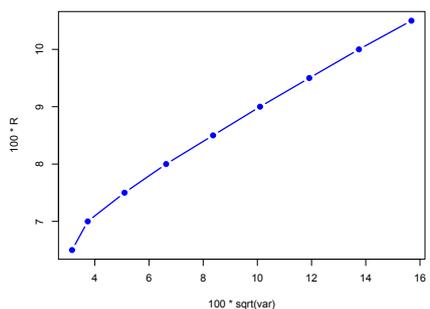
In particolare possiamo facilmente tracciare la frontiera efficiente:

```
> R = seq(0.065,0.105,0.005)
> var = c(0.0010,0.0014,0.0026,0.0044,0.0070,0.0102,0.0142,0.0189,0.0246)
> plot(var,R,type="b",lwd=2,pch=19,col="darkgreen")
```



In verità, ad esser più corretti,

```
> plot(100*sqrt(var),100*R,type="b",lwd=2,pch=19,col="blue")
```



Osserviamo in particolare che a tutti e soli i punti al di sopra - rispettivamente al di sotto - di tale linea corrispondono tutti e soli i portafogli efficienti - rispettivamente inefficienti.

2.3.1 La MVO su Larga Scala: i Problemi della Diversificazione e della Instabilità

Passiamo ora a trattare qualche problema pratico che sorge dall'impiego del modello MVO per la composizione di un portafoglio efficiente $x \in \mathbb{R}^n$ basato su una vasta famiglia di valori, ovvero tale che n risulta oggettivamente piuttosto grande - ad esempio dell'ordine delle centinaia o delle migliaia.

Il punto chiave a riguardo sarebbe quello di riuscire a garantire in un qualche modo una buona diversificazione per il generico portafoglio efficiente, mentre invece non vi è alcuna ragione in generale per aspettarsi una simile proprietà delle soluzioni del modello MVO di Markowitz così com'è.

Abbiamo anzi già anticipato nell'introduzione che al contrario questo modello parrebbe tendere da una parte a produrre portafogli con pesi irragionevolmente consistenti in poche classi di beni, e che d'altra parte questo avverrebbe per di più in un contesto spesso affetto da alta instabilità.

Un'inclinazione del genere viene attribuita anche agli inevitabili errori di stima che affliggerebbero i suoi parametri μ e Σ e che, di conseguenza, avrebbero un effetto macroscopico verosimilmente amplificato proprio dalla loro ingente quantità: pensiamo al fatto che mentre per μ occorrono n stime, dove comunque stiamo immaginando n elevato, per Σ ne servono ben $\frac{n(n-1)}{2} = \mathcal{O}(n^2)$.

Il nostro lavoro volgerà al termine discutendo appunto tutta una serie di ritocchi più o meno sostanziali che potrebbero venire a perfezionare il modello MVO di Markowitz.

Riduzione del Rischio Settoriale nel Modello MVO per la Diversificazione. L'approccio meno sofisticato per spingere il modello a creare portafogli abbastanza diversificati consiste nell'imporre vincoli lineari aggiuntivi sui portafogli ammissibili $x \in \chi$ che viceversa impediscano l'eventualità di composizioni marcatamente sbilanciate: ad esempio potremmo semplicemente imporre un limite superiore uniforme su ogni x_i , $i = 1, \dots, n$, del tipo

$$x_i \leq m \text{ per ogni } i = 1, \dots, n, \text{ o cioè } x \leq m \cdot e,$$

dove naturalmente $m \in (0, 1)$ è una costante pre-fissata - ed $e = e(n) := (1, \dots, 1)^T \in \mathbb{R}^n$.

In modo leggermente più fine ma comunque analogo potremmo ridurre il rischio settoriale raggruppando assieme tutti gli investimenti sui titoli di una data classe di valori e ponendo quindi un limite superiore sull'esposizione a tale classe, e ciò classe dopo classe:

$$\sum_{i \text{ di classe } k} x_i \leq m_k.$$

Tattiche di questo genere portano tuttavia a procedimenti risolutivi che solitamente risultano piuttosto dispendiosi.

Ricambio di Portafoglio e Costi di Transizione nel Modello MVO per l'Instabilità. L'idea è quella di aggiungere un vincolo lineare di ricambio al portafoglio tale per cui il passaggio dalle attuali partecipazioni $x^0 = (x_1^0, \dots, x_n^0)^T \in \mathbb{R}^n$ ad un dato portafoglio desiderato sia forzato a costare una quantità di denaro contante di al più $h > 0$ assegnato.

In parecchie applicazioni pratiche del modello MVO a dimensione alta la matrice Σ delle covarianze risulta infatti quasi-singolare e di conseguenza la decisione ottimale può a maggior ragione cambiare significativamente in corrispondenza a variazioni anche piccole dei dati di input del modello stesso: imponiamo appunto vincoli lineari di ricambio per evitare proprio cambiamenti troppo sproporzionati nel ri-ottimizzare il portafoglio.

Se dunque indichiamo con $y_i \in \mathbb{R}_+$ il valore corrispondente a quanto è stato acquistato dell' i -esimo bene e con $z_i \in \mathbb{R}_+$ il valore corrispondente a quanto invece è stato venduto dell' i -esimo bene, allora i vincoli di ricambio sono

$$\begin{cases} y_i \geq 0 \text{ e } x_i - x_i^0 \leq y_i \text{ per ogni } i = 1, \dots, n \\ z_i \geq 0 \text{ e } x_i^0 - x_i \leq z_i \text{ per ogni } i = 1, \dots, n \\ \sum_{i=1}^n (y_i + z_i) \leq h. \end{cases}$$

A completare il tutto possiamo introdurre direttamente nel modello pure i costi di transizione, almeno quando supponiamo di ri-ottimizzare il portafoglio solo una volta per periodo un portafoglio: se infatti $t_i \in \mathbb{R}_+$ è un costo di transizione proporzionale a y_i e se $t'_i \in \mathbb{R}_+$ è un costo di transizione proporzionale a z_i , allora in definitiva un portafoglio ri-ottimizzato è ottenuto risolvendo $\min_x x^T \Sigma x$, $x \in \mathbb{R}^n$, sui vincoli lineari dati da

$$\begin{cases} \sum_{i=1}^n (\mu_i x_i - t_i y_i - t'_i z_i) \geq R \quad (R > 0 \text{ assegnato}) \\ \sum_{i=1}^n x_i = 1, \quad \sum_{i=1}^n (y_i + z_i) \leq h \\ x_i - x_i^0 \leq y_i \text{ e } x_i^0 - x_i \leq z_i \text{ per ogni } i = 1, \dots, n \\ x_i \geq 0, \quad y_i \geq 0, \quad z_i \geq 0 \text{ per ogni } i = 1, \dots, n. \end{cases}$$

Regressione Lineare Multipla nel CAPM per la Stima dei Parametri del Modello MVO. Sempre a causa della grave instabilità della quale soffre il metodo di MVO riteniamo pressoché ovvio che in generale non converrà prendere come parametri μ e Σ del modello quelli direttamente deducibili dalle varie possibili serie storiche dei passati rendimenti, a differenza di quanto invece abbiamo fatto nel corso dell'Esempio 4. Markowitz stesso suggerisce piuttosto di appoggiarsi al CAPM facendo giocare i beta dei diversi titoli come spieghiamo qui di seguito.

Riprendiamo anzitutto la formula d'equilibrio data dal CAPM a proposito del rendimento atteso di una data attività i ed appliciamola a tutto un insieme di periodi $t = 0, \dots, T$, dove $T \in \mathbb{N} \setminus \{0\}$ è fissato, per i quali sia conosciuto il rendimento lordo $r_{i,t}$ dell'attività i stessa al tempo t : per ogni $t = 0, \dots, T$ otteniamo subito

$$\mathbf{E}[r_{i,t}] = \beta_{i,m;t} (\mathbf{E}[r_{m,t}] - r_{f,t}) + r_{f,t},$$

dove $r_{f,t}$ indica il rendimento certo - ovvero privo di rischio - al tempo t , dove $r_{m,t}$ rappresenta il rendimento di mercato al tempo t e dove $\beta_{i,m;t} := \frac{\text{Cov}[r_{i,t}, r_{m,t}]}{\text{Var}[r_{m,t}]}$ è il coefficiente beta sempre dell'attività i e al tempo t .

L'idea ora è semplicemente quella di eliminare in stima la dipendenza dalla variabile tempo t dai vari coefficienti beta sostituendo ciascun $\beta_{i,m;t}$ con un altro $\tilde{\beta}_{i,m}$ ottenuto procedendo di regressione lineare multipla dal CAPM, per far intervenire infine i due vettori noti $r_f := (r_{f,0}, \dots, r_{f,T})^T$ e $r_m := (r_{m,0}, \dots, r_{m,T})^T$ di \mathbb{R}^{T+1} :

$$r_{i,t} = \tilde{\beta}_{i,m}(r_{m,t} - r_{f,t}) + r_{f,t} + \varepsilon_{i,t},$$

dove per ogni $t = 0, \dots, T$ è stata aggiunta la misura di rischio idiosincratico dell'attività i al tempo t data da $\varepsilon_{i,t}$.

Per ogni attività i , o per ogni coppia di attività diverse i e j , indichiamo con $\varepsilon_i := (\varepsilon_{i,0}, \dots, \varepsilon_{i,T})^T \in \mathbb{R}^{T+1}$, assumiamo quindi che in senso empirico siano $\mathbf{E}[\varepsilon_i] = 0$ e $\mathbf{Cov}[\varepsilon_i, \varepsilon_j] = 0$ - in quanto a ciascun ε_i stiamo attribuendo il ruolo di un errore statistico -, ed indichiamo infine con $\sigma_{\varepsilon_i}^2 := \mathbf{Var}[\varepsilon_i] = \mathbf{Cov}[\varepsilon_i, \varepsilon_i]$.

Determinate così delle stime $\tilde{\beta}_{i,m}$ dei coefficienti $\beta_{i,m;t}$ mediante il classico metodo dei minimi quadrati all'interno del modello regressivo, ricaviamo subito delle stime $\tilde{\mu}_i$ dei μ_i semplicemente passando ai valori attesi empirici:

$$\tilde{\mu}_i := \tilde{\beta}_{i,m}(\mathbf{E}[r_m] - \mathbf{E}[r_f]) + \mathbf{E}[r_f] \text{ per ogni } i.$$

Otteniamo in modo analogo delle stime $\tilde{\sigma}_{i,j}$ dei $\sigma_{i,j} = \mathbf{Cov}[r_i, r_j]$ calcolando tutte le covarianze empiriche ed usando il fatto che il vettore r_f corrisponde in ogni caso ad una variabile reale non-aleatoria, ossia deterministica:

$$\begin{cases} \tilde{\sigma}_{i,i} := \tilde{\beta}_{i,m}^2 \mathbf{Var}[r_m] + \sigma_{\varepsilon_i}^2 \text{ per ogni } i \\ \tilde{\sigma}_{i,j} := \tilde{\beta}_{i,m} \tilde{\beta}_{j,m} \mathbf{Var}[r_m] \text{ per ogni } i \neq j. \end{cases}$$

Peccato tuttavia che tanto questo metodo di regressione lineare, quanto comunque ogni altro metodo di stima presuntamente intelligente dei parametri μ e Σ , non smorzi affatto l'estrema sensibilità della soluzione ottima rispetto ai dati d'input iniziali: per apprezzare questa drammatica problematica risulta generalmente sufficiente anche solo una piccola variazione di uno soltanto dei μ_i .

Una possibile idea per tentare di ovviare a tale problema è stata data da R. O. Michaud a fine anni '90: descritta molto sommariamente, si tratta di un procedimento di ri-campionamento di tutti i vari profitti atto a generare tanti parametri alternativi μ e Σ con lo scopo di risolvere ripetutamente i relativi problemi di MVO e di combinare quindi in un qualche modo i diversi portafogli ottimi così ottenuti.

Un'idea ancora migliore risulta però quella di modificare il modello nella sua essenza.

2.3.2 Il Modello di Black-Litterman per la Stima dei Rendimenti Attesi del Modello MVO

Nel corso dell'introduzione abbiamo affermato che il modello BL arricchisce quello di MVO nel conferire un'importanza fondamentale sia all'equilibrio di mercato che a tutte quei possibili punti di vista che il generico investitore potrebbe assumere nei confronti della complicata evoluzione del mercato stesso.

Con lo scopo di meglio comprendere l'intuizione principale sulla quale poggia l'approccio di F. Black e di R. Litterman, immaginiamoci per un momento di conoscere completamente la reale struttura del rendimento r_i del generico bene i supponendola precisamente come una combinazione lineare di tre fattori speciali:

$$r_i = \pi_i + \gamma_i Z + \nu_i \text{ per ogni } i,$$

dove

- (i) π_i è una variabile deterministica che rappresenta il premio al rischio all'equilibrio;
- (ii) Z è una componente aleatoria comune ad ogni titolo e $\gamma_i \in \mathbb{R}$ misura il suo impatto sul bene i ;
- (iii) ν_i indica infine uno shock aleatorio additivo indipendente da Z .

Assumiamo come nota pure la matrice Σ delle covarianze fra i diversi valori e concentriamoci quindi sul problema del calcolo del vettore μ dei rispettivi rendimenti attesi: dalla precedente relazione strutturale deduciamo che

$$\mu_i = \pi_i + \gamma_i \mathbf{E}[Z] + \mathbf{E}[\nu_i] \text{ per ogni } i.$$

Adesso il punto essenziale è quello di fare una scelta a riguardo del ruolo dei fattori π_i , Z e ν_i non supponendo assolutamente che sia $\mathbf{E}[Z] = \mathbf{E}[\nu_i] = 0$ per ogni i , nonostante certo la semplice struttura in questione corrisponda sostanzialmente ad un legame d'equilibrio.

L'idea è piuttosto quella di ipotizzare che i rendimenti in eccesso previsti siano variabili aleatorie sì inosservabili ma le cui distribuzioni siano univocamente determinate proprio dai soli scalari $\mathbf{E}[Z]$ e $\mathbf{E}[\nu_i]$, al variare di i , anziché dalle medesime variabili Z e ν_i considerate nella loro intera complessità.

Un'ipotesi del genere implicherà che la matrice delle covarianze sia in realtà "più ridotta" di quanto potesse sembrare in un primo momento, per così dire: assumiamo di poterla sostituire con

$$\tau \Sigma, \text{ dove } \tau \in (0, 1].$$

Mentre tutte queste informazioni possono esser immaginate come note a ciascuno dei generici investitori, veniamo finalmente a differenziare le possibili posizioni d'investimento aggiungendo diverse previsioni del tutto soggettive a riguardo dei profitti futuri, delle quali l'espressione tipica avrà una forma analoga alla seguente: preso $\eta \in [0, 100]$,

«mi aspetto che a fine periodo il rendimento del bene A surclassi il rendimento del bene B di un $\eta\%$ ».

Se dunque affianchiamo ad ogni simile previsione pure una misura della fiducia ad essa relativa, allora possiamo incorporare matematicamente queste previsioni direttamente nei profitti r_i d'equilibrio a generare così delle specie di leggi condizionali - dalle quali poter calcolare in particolare il vettore μ .

Tornando a rinominare con $n \in \mathbb{N} \setminus \{0\}$ il numero complessivo dei beni, ossia quello tale che $\mu \in \mathbb{R}^n$ e $\Sigma \in \mathbb{R}^{n \times n}$, e posti $r := (r_1, \dots, r_n)^T \in \mathbb{R}^n$ e $\pi := (\pi_1, \dots, \pi_n)^T \in \mathbb{R}^n$, secondo le ipotesi tecniche di Black e Litterman il vettore r è un vettore aleatorio avente distribuzione che sia data dal prodotto di due leggi normali n -variate, ossia delle leggi di due vettori gaussiani in \mathbb{R}^n - aventi sempre ottime proprietà computazionali -, sostanzialmente corrispondenti a due opportune regressioni lineari multiple fatte in sequenza e comunque sommariamente descritte qui di seguito.

(*) La prima legge coincide con $\mathcal{N}_n(\pi, \tau\Sigma)$ e come tale rappresenta quindi il profitto all'equilibrio di mercato. Tipicamente viene scelto un $\tau \in (0, 1)$ piccolo, ovvero con $\tau \ll 1$, anche dovutamente alla constatazione empirica del fatto che la varianza di ciascun μ_i risulta di solito notevolmente inferiore rispetto alla varianza σ_i^2 del profitto dell' i -esimo bene. Notiamo che in ogni caso $\tau\Sigma$ è definita positiva quando pure Σ lo è.

(**) La seconda legge ingloba invece l'opinione complessiva su μ che l'investitore ha e alla quale egli vuole effettivamente dare un peso in base alla sua incertezza a riguardo. Formalmente quest'opinione entra in gioco imponendo che μ sia a sua volta un vettore aleatorio vincolato al seguente legame lineare multi-dimensionale:

$$P\mu = q + \varepsilon$$

dove son tutti dati forniti appunto dall'investitore $k \in \mathbb{N} \setminus \{0\}$, $P \in \mathbb{R}^{k \times n}$, $q \in \mathbb{R}^k$ e $\varepsilon \sim \mathcal{N}_k(0, \Omega)$ con $\Omega \in \mathbb{R}^{k \times k}$ matrice diagonale invertibile, ed in particolare ε è un vettore gaussiano in \mathbb{R}^k di componenti mutuamente indipendenti. In definitiva sussiste la relazione $P\mu - q \sim \mathcal{N}_k(0, \Omega)$, o equivalentemente

$$P\mu \sim \mathcal{N}_k(q, \Omega) .$$

È possibile verificare che la risultante legge condizionale di μ resta normale n -variata e di media data da

$$\bar{\mu} := [(\tau\Sigma)^{-1} + P^T\Omega^{-1}P]^{-1}[(\tau\Sigma)^{-1}\pi + P^T\Omega^{-1}q] :$$

è esattamente questa che Black e Litterman usano come stima del parametro μ per il modello MVO di Markowitz.

Nella pratica l'investitore in questione ha $k \in \mathbb{N} \setminus \{0\}$ previsioni sui profitti futuri aventi valor medio $q \in \mathbb{R}^k$ e rispettivi livelli di fiducia $\omega_h := \Omega_{h,h}$, con $h = 1, \dots, k$, nel mentre che $P \in \mathbb{R}^{k \times k}$ dirige il tutto incrociando opportunamente le componenti di μ alle quali le suddette previsioni vogliono riferirsi.

Tutto questo meriterebbe senz'altro un accurato approfondimento, però dobbiamo accontentarci solo di capire un po' meglio la questione riprendendo l'Esempio 4 ed inventandoci almeno un paio di previsioni più o meno plausibili su stocks, bonds oppure MM per applicare quindi la precedente formula di Black-Litterman d'approssimazione del parametro μ e ri-ottimizzare infine il portafoglio all'interno del modello MVO corrispondente.

Esempio 5. Cominciamo ricordando anzitutto quelle due tabelle alle quali nell'Esempio 4 abbiamo fatto corrispondere i due parametri μ e Σ , ovvero rispettivamente

	Stocks	Bonds	MM			Stocks	Bonds	MM
<i>Media Geometrica</i>	10.72%	7.3%	6.27%	e	Stocks	0.02839	0.00388	0.00021
					Bonds	0.00388	0.01148	-0.00024
					MM	0.00021	-0.00024	0.00118

Per determinare la prima legge normale 3-variata, poniamo ad esempio $\tau := 0.1$ e soprattutto

$$\pi := \begin{bmatrix} 0.1072 \\ 0.073 \\ 0.0627 \end{bmatrix} \quad e \quad \Sigma := \begin{bmatrix} 0.02839 & 0.00388 & 0.00021 \\ 0.00388 & 0.01148 & -0.00024 \\ 0.00021 & -0.00024 & 0.00118 \end{bmatrix} .$$

Per determinare invece la seconda legge 3-variata, immaginiamo intanto i due seguenti semplici scenari - per cui $k := 2$:

- una previsione piuttosto forte sull'eventualità che a fine periodo, poniamo tra un anno, il tasso del MM scenda al 2%;
- una previsione ben più incerta sull'eventualità che siano messi fuori dei BTP - Buoni del Tesoro Poliennali - a scadenza di dieci anni al 5%.

Formalizzando mediante anche il simbolo $\mu := (\mu_S, \mu_B, \mu_M)^T \in \mathbb{R}^3$, stiamo semplicemente assumendo questo:

$$\begin{cases} \mu_M = 0.02 & \text{con un'incertezza - ad esempio - di } \omega_1 = \Omega_{1,1} := 0.00001 \\ \mu_S - \mu_B = 0.05 & \text{con un'incertezza - ad esempio - di } \omega_2 = \Omega_{2,2} := 0.001 . \end{cases}$$

Pertanto possiamo porre infine

$$P := \begin{bmatrix} 0 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix}, \quad q := \begin{bmatrix} 0.02 \\ 0.05 \end{bmatrix} \quad e \quad \Omega := \begin{bmatrix} 0.00001 & 0 \\ 0 & 0.001 \end{bmatrix} .$$

Abbiamo così tutti gli ingredienti necessari e sufficienti della formula di Black-Litterman per la stima del parametro μ :

$$\bar{\mu} := [(\tau\Sigma)^{-1} + P^T\Omega^{-1}P]^{-1}[(\tau\Sigma)^{-1}\pi + P^T\Omega^{-1}q].$$

Ciò che viene trovato dagli autori è riportato nella seguente tabella:

	Stocks	Bonds	MM
$\bar{\mu}$	11.77%	7.51%	2.34%

Notiamo in particolare quanto le nostre previsioni abbiano influenzato la componente del profitto atteso relativa al MM.

In conclusione il problema assegnato diventa un problema di portafoglio efficiente su \mathbb{R}^3 per mezzo del seguente modello MVO di Markowitz avente come parametro delle covarianze sempre Σ mentre come parametro dei valori attesi $\bar{\mu}$ anziché μ :

$$\begin{cases} \min_x x^T \Sigma x \\ \bar{\mu}^T x \geq R \\ e^T x = 1 \\ x \geq 0 \end{cases}$$

dove ancora $x := (x_S, x_B, x_M)^T \in \mathbb{R}^3$ ed $e := (1, 1, 1)^T \in \mathbb{R}^3$, o quindi più esplicitamente

$$\begin{cases} \min_{(x_S, x_B, x_M)} 0.02839x_S^2 + 0.01148x_B^2 + 0.00118x_M^2 + \\ \quad + 2 \times 0.00388x_Sx_B + 2 \times 0.00021x_Sx_M - 2 \times 0.00024x_Bx_M \\ 0.1177x_S + 0.0751x_B + 0.0234x_M \geq R \\ x_S + x_B + x_M = 1 \\ x_S \geq 0, x_B \geq 0, x_M \geq 0, \end{cases}$$

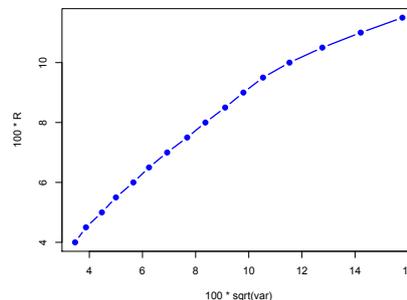
al variare ad esempio di $R > 0$ tra 2.34% \rightsquigarrow 4% e 11.77% \rightsquigarrow 11.5% con incrementi dello 0.5%.

Gli autori risolvono questo problema di QP sempre grazie all'ausilio di un software ed arrivano a questi risultati:

R	Varianza ottima	Stocks	Bonds	MM
0.040	0.0012	0.08	0.17	0.75
0.045	0.0015	0.11	0.21	0.68
0.050	0.0020	0.15	0.24	0.61
0.055	0.0025	0.18	0.28	0.54
0.060	0.0032	0.22	0.31	0.47
0.065	0.0039	0.25	0.35	0.40
0.070	0.0048	0.28	0.39	0.33
0.075	0.0059	0.32	0.42	0.26
0.080	0.0070	0.35	0.46	0.19
0.085	0.0083	0.38	0.49	0.13
0.090	0.0096	0.42	0.53	0.05
0.095	0.0111	0.47	0.53	0
0.100	0.0133	0.58	0.42	0
0.105	0.0163	0.70	0.30	0
0.110	0.0202	0.82	0.18	0
0.115	0.0249	0.94	0.06	0

Disegniamo in particolare la frontiera efficiente:

```
> R = seq(0.04, 0.115, 0.005)
> var = c(0.0012, 0.0015, 0.0020, 0.0025, 0.0032, 0.0039, 0.0048, 0.0059, 0.0070, 0.0083, 0.0096, 0.0111, 0.0133,
+ 0.0163, 0.0202, 0.0249)
> plot(100*sqrt(var), 100*R, type="b", lwd=2, pch=19, col="blue")
```



2.3.3 Il Modello di Konno-Yamazaki: Minimizzazione del Modulo della Deviazione dalla Media

Per questa sotto-sezione potremmo finalmente far intervenire quella parte svolta all'inizio sulla programmazione lineare, poiché infatti H. Konno e H. Yamazaki propongono per l'ottimizzazione di portafoglio un modello di LP anziché uno di vera e propria QP, e lo fanno basandosi sull'osservazione del fatto che due differenti misure di rischio potrebbero esser strettamente legate ed ugualmente appropriate per la desiderata ottimizzazione.

Spieghiamoci meglio venendo subito alla matematica della questione: quali che siano $x = (x_1, \dots, x_n)^T \in \chi$ un portafoglio ammissibile, $r = (r_1, \dots, r_n)^T \in \mathbb{R}^n$ il rendimento aleatorio effettivo e $\mu = (\mu_1, \dots, \mu_n)^T := \mathbf{E}[r] \in \mathbb{R}^n$ il conseguente rendimento atteso, calcoliamo intanto quelle due misure di rischio associate alla strategia x chiamate

--> la *volatilità* del profitto di x , e data da $\sigma = \sigma(x) := \mathbf{E}[\sum_{i=1}^n (r_i - \mu_i)x_i]^2]^{\frac{1}{2}}$ (“ \cong ” $\sqrt{\mathbf{Var}[x]}$), e

--> la L_1 -*risk* del profitto di x , e data da $w = w(x) := \mathbf{E}[\sum_{i=1}^n (r_i - \mu_i)x_i]$,

per le quali notiamo in particolare che $(r_i - \mu_i)x_i = x_i r_i - \mathbf{E}[x_i r_i]$ per ogni $i = 1, \dots, n$.

Allora risultano proprio σ e w ad esser talvolta strettamente legate, e ciò in un senso precisato dal seguente risultato di probabilità elementare dovuto sempre a Konno e Yamazaki - del quale omettiamo la semplice dimostrazione.

Teorema (Konno-Yamazaki). *Qualora r fosse un vettore gaussiano normale in \mathbb{R}^n , ovvero tale che $r \sim \mathcal{N}_n(0, I_n)$, allora varrebbe il seguente legame lineare tra $w(\cdot)$ e $\sigma(\cdot)$: per ogni $x \in \chi$,*

$$w(x) = \sqrt{\frac{2}{\pi}} \sigma(x).$$

Il nostro interesse nei confronti della precedente proposizione è giustificato da un suo immediato corollario: se $r \sim \mathcal{N}_n(0, I_n)$, allora il problema di minimizzare $\mathbf{Var}[\cdot]$ su χ è del tutto equivalente a quello di minimizzare sempre su χ giustappunto il modulo della deviazione dalla media $w(\cdot)$: così il problema di MVO può assumere la forma

$$\begin{cases} \min_x \mathbf{E}[\sum_{i=1}^n (r_i - \mu_i)x_i] \\ \mu^T x \geq R \\ e^T x = 1 \\ 0 \leq x \leq m \end{cases}$$

dove al solito $R > 0$ e $m \in [0, 1]^n$ sono costanti fissate - ed $e = e(n) := (1, \dots, 1) \in \mathbb{R}^n$.

Il modello di Konno-Yamazaki, chiamato anche *modello di mean-absolute deviation* o brevemente *modello MAD*, indica proprio un problema di minimizzazione della misura L_1 -*risk* $w(\cdot)$ di una forma simile. Prestiamo attenzione al fatto che generalmente il vettore r non sarà gaussiano e tanto meno gaussiano standard, ma che tuttavia il modello MAD riesce spesso a produrre portafogli efficienti per la misura L_1 -*risk* che risultano composti in modo del tutto confrontabile rispetto a quelli efficienti nel senso di Markowitz.

Vogliamo quindi riprendere per un'ultima volta l'Esempio 4 proprio per constatare una tale tipica confrontabilità di composizione, e di conseguenza converrà dapprima linearizzare esplicitamente questo modello MAD - grazie a variabili addizionali ausiliarie - per poterlo poi trattare come problema di LP.

Dato $T \in \mathbb{N} \setminus \{0\}$ ed al variare di $i = 1, \dots, n$ e $t = 1, \dots, T$, supponiamo dunque di aver a disposizione sotto forma di dati storici, od eventualmente di previsioni future verosimili, la realizzazione $r_{i,t}$ di r_i al tempo t : allora $\mu_i = \mathbf{E}[r_i] = \frac{1}{T} \sum_{t=1}^T r_{i,t}$ per ogni $i = 1, \dots, n$ ed analogamente è davvero facile rendersi conto che

$$w = w(x) = \mathbf{E}[\sum_{i=1}^n (r_i - \mu_i)x_i] = \frac{1}{T} \sum_{t=1}^T \left| \sum_{i=1}^n (r_{i,t} - \mu_i)x_i \right|, \quad x \in \chi.$$

Grazie ora ad un elementare cambio di variabile del tipo $|s| = y + z$ dove $y \geq 0$ e $z \geq 0$ sono tali che $s = y - z$, com'è ad esempio per la coppia $y = s^+ := \max\{s, 0\}$ e $z = s^- := \min\{s, 0\}$, nel tendere all'obiettivo di minimizzare appunto $y + z$ accadrà che al più uno soltanto fra y e z resti positivo mentre l'altro diverrà automaticamente zero: possiamo finalmente riscrivere il precedente modello MAD come il problema di LP su $\mathbb{R}_{(x,y,z)}^{n+2T}$ della forma

$$\begin{cases} \min_{x,y,z} \sum_{t=1}^T (y_t + z_t) \\ y_t - z_t = \sum_{i=1}^n (r_{i,t} - \mu_i)x_i \text{ per ogni } t = 1, \dots, T \\ \mu^T x \geq R \\ e^T x = 1 \\ 0 \leq x \leq m \\ y_t \geq 0 \text{ e } z_t \geq 0 \text{ per ogni } t = 1, \dots, T. \end{cases}$$

Esempio 6. *Gli autori risolvono il problema di portafoglio basato su stocks, bonds e MM dell'Esempio 4 adattandolo appunto al modello MAD in questa sua formulazione come problema di LP: fissato per l'esattezza $m := e = (1, 1, 1)^T$ e facendo correre $R > 0$ dal 6.5% al 10.5% coi soliti incrementi dello 0.5%, quanto ottengono è dato da*

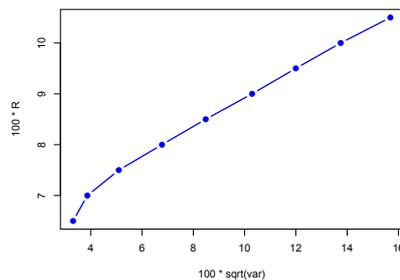
R	Varianza ottima	Stocks	Bonds	MM
0.065	0.0011	0.05	0.01	0.94
0.070	0.0015	0.15	0.04	0.81
0.075	0.0026	0.25	0.11	0.64
0.080	0.0046	0.32	0.28	0.40
0.085	0.0072	0.42	0.32	0.26
0.090	0.0106	0.52	0.37	0.11
0.095	0.0144	0.63	0.37	0
0.100	0.0189	0.78	0.22	0
0.105	0.0246	0.93	0.07	0

Ecco in particolare la frontiera efficiente:

```

> R = seq(0.065,0.105,0.005)
> var = c(0.0011,0.0015,0.0026,0.0046,0.0072,0.0106,0.0144,0.0189,0.0246)
> plot(100*sqrt(var),100*R,type="b",lwd=2,pch=19,col="blue")

```



2.3.4 Massimizzazione dell'Indice di Sharpe

Proponiamo infine una modellizzazione di un problema di portafoglio ottimo ancora come un problema quasi di QP, ma non esattamente sotto forma di MVO, la quale formalizzi una ben precisa risoluzione geometrica del problema stesso ottenuta in modo comunque altamente legato alla frontiera efficiente nel senso di Markowitz.

Torniamo per questo a considerare quel $x_{min} \in \chi$ tale che $\mathbf{Var}[x_{min}] \leq \mathbf{Var}[x]$ per ogni $x \in \chi$, quindi $R_{min} = \mu^T x_{min} = \mathbf{E}[x_{min}]$, nonché pure il massimo valore atteso $R_{max} \in (0, \infty)$ fra quelli dei portafogli ammissibili. Al variare dunque di $R > 0$ in $[R_{min}, R_{max}]$, se $x_R \in \chi$ è la soluzione del solito problema di MVO

$$\begin{cases} \min_x x^T \Sigma x \\ \mu^T x \geq R \\ Ax = b \\ Cx \geq d \end{cases}$$

e se di nuovo $\sigma(R) : [R_{min}, R_{max}] \rightarrow \mathbb{R}$, $\sigma(R) = \sqrt{\mathbf{Var}[x_R]} = \sqrt{x_R^T \Sigma x_R}$, allora riscriviamo la frontiera efficiente di χ come il grafico di σ - nello spazio $[R_{min}, R_{max}] \times \mathbb{R}_+$ -

$$E := \{(R, \sigma(R)) = (R, \sqrt{\mathbf{Var}[x_R]}) : R \in [R_{min}, R_{max}]\}.$$

All'interno della classe dei valori sopra i quali organizziamo il portafoglio, introduciamo adesso un titolo privo di rischio e di profitto $r_f \geq 0$ per il quale assumiamo che valga $r_f < R_{min}$, o equivalentemente $r_f < R$ per ciascun $R \in [R_{min}, R_{max}]$: infatti a $x_{R_{min}}$ resta associato un rischio > 0 e per ciò è perfettamente naturale che esso presenti un rendimento atteso più alto rispetto ad un qualsiasi dato rendimento certo.

Ebbene, tutti i diversi profili sia del profitto ch  del rischio relativi alle possibili combinazioni di un portafoglio rischioso con un tale bene r_f possono esser rappresentati graficamente come una semiretta in \mathbb{R}_+^2 , ove conveniamo di porre in ascissa la deviazione standard ed in ordinata la media, la quale si sviluppa a partire dal punto $(0, r_f)$ e la quale   chiamata quindi *linea di allocazione del capitale*, di acronimo inglese *CAL*.

L'efficienza della strategia di portafoglio corrispondente ad una certa prescelta CAL pu  esser tradotta in modo equivalente mediante la propriet  geometrica di tale retta di giacere strettamente al di sopra della frontiera efficiente E di χ , e cos  la CAL che pu  esser considerata come ottimale fra tutte risulta quella che coincide esattamente con la retta tangente ad E che ha principio in $(0, r_f)$: il punto di tangenza fra queste due linee rappresenta proprio quel portafoglio avente rischio ottimale.

In altri termini abbiamo che la CAL ottima coincide con quella che presenta la pendenza minima fra quelle di tutte le altre CALs "efficienti" che partono da $(0, r_f)$, pendenza che poi   un sotto-differenziale della funzione $\sigma(\cdot)$.

Possiamo descrivere matematicamente tutto questo per mezzo del processo di ricerca di quel portafoglio $x^* \in \chi$ che massimizza appunto l'*indice o rapporto di Sharpe*, definito ponendo

$$h(x) := \frac{\mu^T x - r_f}{(x^T \Sigma x)^{\frac{1}{2}}}, x \in \mathbb{R}^n \setminus \{0\}.$$

Pertanto il problema di portafoglio rischioso ottimo pu  assumere attraverso $h(\cdot)$ la seguente forma:

$$\begin{cases} \max_x \frac{\mu^T x - r_f}{\sqrt{x^T \Sigma x}} \\ Ax = b \\ Cx \geq d. \end{cases}$$

Tuttavia avvistiamo subito una preoccupante difficolt  conseguente a questa formulazione: mentre la regione ammissibile resta poliedrale, la funzione obiettivo $h(x)$   assai complicata e soprattutto possibilmente non-concava.

Concludiamo la sezione presentando quindi un metodo diretto che costruisce esplicitamente un problema quasi di QP equivalente al precedente sotto un paio di ragionevolissime ulteriori assunzioni: che valga $e^T x = 1$ per ogni portafoglio ammissibile $x \in \chi$ - $e = (1, \dots, 1)^T \in \mathbb{R}^n$ - e che fra questi ne esista uno speciale $\hat{x} \in \chi$ avente $\mu^T \hat{x} > r_f$.

In effetti, se fosse $\mu^T x \leq r_f$ per ogni $x \in \chi$, allora non vi sarebbe alcunché da ottimizzare in quanto l'investimento a rischio zero dominerebbe tutti gli altri in modo davvero banale.

Proposizione. *Supponiamo che risulti $e^T x = \sum_{i=1}^n x_i = 1$ per ogni $x \in \chi$ e che esista $\hat{x} \in \chi$ con $\mu^T \hat{x} - r_f > 0$. Allora esiste uno ed un solo $x^* \in \chi$ che massimizzi l'indice di Sharpe $h(\cdot)$ su χ , e può essere determinato come segue: definito*

$$\chi_+ := \{(y, k) \in \mathbb{R}^n \times (0, \infty) : \frac{y}{k} \in \chi\} \cup \{(0, 0)\}$$

e considerata la soluzione $(y^*, k^*) \in \chi_+$ del problema di programmazione quadratica convessa su $\mathbb{R}^{n+1}_{(y,k)}$

$$\begin{cases} \min_{y,k} y^T \Sigma y \\ (y, k) \in \chi_+ \\ (\mu - r_f e)^T y = \sum_{i=1}^n (\mu_i - r_f) y_i = 1, \end{cases}$$

allora è $x^* = \frac{y^*}{k^*}$.

Dimostrazione. Anzitutto χ_+ è non vuoto, ad esempio perché palesemente $\chi \times \{1\} \subseteq \chi_+$, ed è chiaramente un cono: per ogni $z \in \chi_+$ e per ogni $\lambda \geq 0$, anche $\lambda z \in \chi_+$. Si tratta in verità di un insieme che è pure convesso, visto che

$$\chi = \{x \in \mathbb{R}^n : Ax = b, Cx \geq d\} \Rightarrow \chi_+ = \{(y, k) \in \mathbb{R}^n \times \mathbb{R} : k > 0, Ay - bk = 0, Cy - dk \geq 0\} \cup \{(0, 0)\}.$$

Inoltre la chiusura di χ e l'inclusione dello $(0, 0) \in \mathbb{R}^n \times \mathbb{R}$ fanno di χ_+ un sottoinsieme di \mathbb{R}^{n+1} chiuso a sua volta: ossia in definitiva un cono, convesso, chiuso e non vuoto di \mathbb{R}^{n+1} .

Passiamo quindi al resto cominciando con l'osservare da una parte che, per ogni $x \in \chi$, la prima ipotesi $e^T x = \sum_{i=1}^n x_i = 1$ implica subito che $\mu^T x - r_f = (\mu - r_f e)^T x$ e che, in particolare, $h(x) = \frac{(\mu - r_f e)^T x}{(x^T \Sigma x)^{\frac{1}{2}}}$. D'altra parte poi che la seconda ipotesi $\mu^T \hat{x} - r_f > 0$ equivale a $h(\hat{x}) > 0$ e che, di conseguenza, possiamo restringerci senz'altro ai soli $x \in \chi$ che presentino a loro volta la proprietà $\mu^T x - r_f > 0$.

A questo punto leghiamo i due problemi d'ottimizzazione per mezzo del seguente elementare cambio di variabili:

$$\begin{cases} k \doteq \frac{1}{\mu^T x - r_f} = \frac{1}{(\mu - r_f e)^T x} \\ y \doteq kx. \end{cases}$$

Infatti, per ogni $x \in \chi$ tale che appunto $\mu^T x - r_f > 0$, abbiamo chiaramente che $k \in (0, \infty)$ e che $y \in \mathbb{R}^n$ soddisfa a $\frac{y}{k} \in \chi$, o cioè che insomma $(y, k) \in \chi_+$, ed inoltre che sussiste $(\mu - r_f e)^T y = 1$ in modo davvero ovvio.

Viceversa, per ogni $(y, k) \in \chi_+$ con $(\mu - r_f e)^T y = 1$, quel $x \in \chi$ univocamente determinato come $x = \frac{y}{k}$ è tale per cui $1 = (\mu - r_f e)^T y = (\mu - r_f e)^T kx$ da cui immediatamente l'espressione di k e quindi quella di y .

Verifichiamo infine che cosa accade alle due funzioni obiettivo attraverso questa corrispondenza biunivoca fra gli $x \in \chi$ con $\mu^T x > r_f$ e gli $(y, k) \in \chi_+$ con $(\mu - r_f e)^T y = 1$: notato che $x^T \Sigma x = \frac{1}{k^2} y^T \Sigma y$, o equivalentemente che $(x^T \Sigma x)^{\frac{1}{2}} = \frac{1}{k} (y^T \Sigma y)^{\frac{1}{2}}$, abbiamo in definitiva che

$$\frac{\mu^T x - r_f}{\sqrt{x^T \Sigma x}} = \frac{1}{\sqrt{y^T \Sigma y}}.$$

Con tutto questo la tesi resta dimostrata. □

Riferimenti bibliografici

- [1] Björk, T. (2009), *Arbitrage Theory in Continuous Time*, Oxford University Press.
- [2] Cornuejols, G. - Tütüncü, R. (2006), *Optimization Methods in Finance*, Carnegie Mellon University, Pittsburgh, PA 15213 USA.
- [3] Hull, J.C. (2014), *Options, Futures, and Other Derivatives*, Pearson.
- [4] Lamberton, D. - Lapeyre, B. (2007), *Introduction to Stochastic Calculus Applied to Finance*, Chapman and Hall, CRC Press.

