Università di Pisa
Dipartimento di Informatica

Technical Report: TR-10-09

Introduzione alla System Dynamics

Lorenzo Cioni

Maggio 6, 2010

ADDRESS: Largo B. Pontecorvo 3, 56127 Pisa, Italy. TEL: +39 050 2212700 FAX: +39 050 2212726

Introduzione alla System Dynamics

Lorenzo Cioni

Maggio 6, 2010

Dipartimento di Informatica, Università di Pisa

e-mail: lcioni@di.unipi.it

Abstract

Le presenti note rappresentano un ampliamento e un approfondimento di alcuni degli argomenti svolti dal Prof. Gallo per il Corso di "Modellistica Ambientale" da lui tenuto nel secondo semestre degli Anni Accademici 2005/2006, 2006/2007, 2007/2008, 2008/2009 e 2009/2010 per gli studenti e le studentesse del Corso di Laurea in "Scienze e Tecnologie per l'Ambiente".
Gli appunti sono focalizzati sui concetti base della System Dynamics, con particolare riferimento ai diagrammi ad anelli causali e ai diagrammi di flusso, e contengono esempi di costruzione di modelli. Segnalazioni di errori, inesattezze e lacune sono molto gradite.

Contents

List of Figures

1 Introduzione

I concetti di base che saranno presentati in questa Introduzione sono[1]:

1. il concetto di **modello**;
2. il concetto di **simulazione**;

in un'ottica di tipo **sistemico**.
Con il termine **modello** si definisce una rappresentazione semplificata mentre la **simulazione eseguita** sul modello mira a riprodurre comportamenti e/o elementi significativi dell'oggetto simulato. La relazione fra modello e simulazione è volutamente mantenuta vaga dal momento che i modelli possono essere sia rappresentazioni in scala di porzioni della realtà (è il caso di modelli di navi, di veicoli terrestri e di aerei) sia costruzioni concettuali astratte descritte per mezzo di equazioni matematiche o di programmi per computer.
In entrambi i casi i modelli astraggono da caratteristiche inessenziali agli scopi per i quali un modello è concepito e realizzato e tengono conto solo di grandezze significative per tali scopi. Nel caso di un modello in scala di un'automobile da usare nella galleria del vento, ad esempio, il modello trascura completamente la struttura interna del veicolo e si concentra sulla struttura esterna della quale si vuole studiare il comportamento per quanto riguarda la resistenza alla penetrazione nell'aria e le proprietà aerodinamiche. Nel caso di un modello matematico si compie un passo di astrazione ulteriore dal momento che si definiscono simboli matematici ed equazioni per rappresentare le relazioni fra porzioni di un sistema.
A questo punto risulta necessario introdurre il concetto di **sistema**.

Definizione 1.1 *Un **sistema** è un insieme di elementi **interconnessi** e **interagenti** che agiscono come un tutto per un qualche **scopo**.*

L'uso del termine scopo esclude ogni teleologia o finalismo. Nel caso di sistemi artificiali va inteso come "realizzazione delle specifiche di progetto" mentre in quelli naturali lo si può interpretare come "realizzazione di un punto di vista" e, in quanto tale, non presente nel sistema stesso ma, piuttosto, presente nella mente degli osservatori. La pluralità degli osservatori introduce sia una pluralità di punti di vista sia una pluralità di scopi.

Esempio 1.1 *Esempi di sistemi "pezzi di realtà" di cui si possono concepire i modelli descrittivi sono:*

1. *una città con la dinamica di opportunità di lavoro, di commercio e abitative;*
2. *una ditta con la sua produzione, le vendite, i rapporti con la forza lavoro e i competitori;*

[1] Il contenuto delle sezioni iniziali deriva quasi per intero da [RAD+83].

3. *una linea aerea o marittima con tutti gli elementi necessari per il trasporto di merci e passeggeri.*

Dal punto di vista della Definizione 1.1, sia il modello che l'oggetto rappresentato dal modello sono sistemi. La motivazione del passaggio dal rappresentato al rappresentante è essenzialmente la maggiore semplicità concettuale dei modelli rispetto a ciò che rappresentano. Il motivo base di tale maggiore semplicità risiede nel fatto che un modello è un artefatto mentre il rappresentato è, spesso, la "realtà" (o una sua parte) con tutta la sua **complessità**[2].
Premesso che prenderemo in considerazione solo modelli concettuali, il metodo che seguiremo per la costruzione dei modelli è quello che prevede l'uso di:

1. un approccio sistemico;

2. relazioni dinamiche con retroazione (o feedback).

L'approccio sistemico allo studio dei sistemi enfatizza la connessione fra le varie parti del sistema ovvero enfatizza:

1. la connettività;

2. l'interezza;

secondo una visione intrinsecamente multidisciplinare. Con riferimento all'esempio 1.1 si ha:

1. l'analisi di una città coinvolge discipline come Sociologia, Economia, Geografia fra le altre;

2. l'analisi di una ditta coinvolge discipline come l'Economia dei mercati, la Sociologia dell'Impresa e le varie tecniche di produzione.

L'approccio sistemico si può concretizzare in vari modi, uno dei quali è rappresentato dalla System Dynamics (o Dinamica dei Sistemi) a cui queste note sono, in larga parte, dedicate.

Nota 1.1 *L'approccio che seguiremo può essere definito come un approccio* ***olistico*** *dal momento che considereremo imprescindibili le interrelazioni funzionali fra le parti di un sistema, visto come qualcosa di più della semplice somma delle parti costituenti. L'approccio duale è il cosiddetto approccio* ***riduzionistico*** *che si propone di ricondurre tutte le spiegazioni a quelle delle discipline di base come la fisica in modo che, ad esempio, l'unica differenza fra il vivente e la materia inerte sia una differenza di scala di complessità e il tutto si possa vedere come una semplice somma delle sue parti.*

[2]Il termine **complessità** non sarà definito formalmente. In modo informale si può dire che la complessità è in corrispondenza biunivoca con la quantità di informazione (e di tempo) di cui è necessario disporre per descrivere un sistema nella sua struttura e nel suo comportamento. In tal modo la complessità dipende anche dal livello di astrazione a cui si descrive un sistema.

1.1 Il processo di costruzione del modello

La costruzione di un modello richiede che vengano compiuti alcuni passi preliminari quali:

1. il riconoscimento che un problema complesso non può essere risolto senza una analisi approfondita del problema stesso e una profonda conoscenza delle discipline in esso coinvolte;

2. la definizione di un metodo per strutturare la conoscenza del problema per poterlo analizzare individuando quali fattori inserire e quali escludere dal modello e come rappresentare tali fattori (usando le cosiddette metafore, vedi oltre);

3. la definizione di strumenti che consentano di tenere traccia di tutte le interrelazioni che sono state individuate fra gli elementi costitutivi di un sistema.

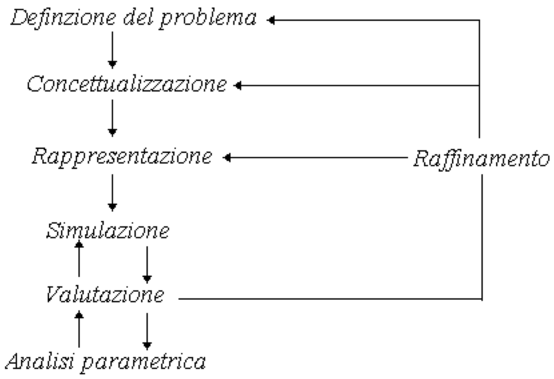

Figure 1: *Il processo di costruzione del modello (dettagli nel testo)*

In merito all'ultimo punto le relazioni devono essere quantificate, ovvero tradotte in relazioni di tipo quantitativo, in modo da arrivare alla definizione di un **modello di simulazione**. Il modello di simulazione è un modello di cui si può studiare l'evoluzione nel tempo e determinare come le grandezze rappresentative dei suoi elementi costitutivi variano nel tempo a partire da certe condizioni iniziali che, in genere, individuano una condizione di **equilibrio dinamico** del sistema[3].
Con riferimento alla figura 1, il processo di definizione ed utilizzo di un modello di simulazione attraverso l'approccio sistemico si compone dei passi (o fasi) seguenti.

[3]I concetti introdotti saranno chiariti nelle sezioni che seguono.

1. La prima fase consiste nella individuazione e definizione di un problema che si vuole analizzare usando la System Dynamics (SD). Perchè ciò sia possibile il problema deve essere intrinsecamente dinamico ovvero deve essere caratterizzabile da grandezze che variano nel tempo le cui interrelazioni siano descrivibili mediante relazioni di **causa÷effetto** che definiscono anelli di retroazione (o di feedback).

2. Nella seconda fase si passa alla individuazione delle grandezze significative che si vogliono includere nel modello, ad una loro **rappresentazione** su carta e alla individuazione dei legami reciproci (vedi la sezione 1.2).

3. Come terzo passo si devono quantificare le relazioni esistenti fra le grandezze suddette mediante equazioni che permettono di caratterizzare l'evoluzione nel tempo, su un orizzonte di tempo specificato e con una cadenza specificata[4], delle grandezze caratteristiche del modello e, quindi, del modello stesso.

4. La fase successiva consiste nel simulare il modello ovvero, dati i valori iniziali delle varie grandezze, nel calcolare i valori delle grandezze ogni Δt su tutto T_∞.

5. Il passo successivo consiste nel valutare se il modello mostra il comportamento atteso. La valutazione può essere di due tipi:

   (a) qualitativo qualora si possano valutare solo gli andamenti delle variabili (o grandezze) per vedere se rispecchiano quelli attesi o se vi si discostano, in che misura e in che modo;

---

[4] Intuitivamente questo significa che:

(a) si stabilisce un istante di inizio t_0 e uno di fine t_∞ del processo di simulazione (che definiscono l'orizzonte temporale T_∞);

(b) si stabilisce ogni quanto tempo Δt, fra tali istanti, si valutano i valori successivi delle variabili.

Si può intuire che:

(a) tanto più ampio è l'orizzonte temporale quanto più computazionalmente pesante è la valutazione delle equazioni descrittive del modello;

(b) l'orizzonte temporale deve essere fissato ad un valore che consenta di apprezzare il comportamento del sistema ovvero consenta al sistema di passare da uno stato iniziale da noi ipotizzato ad uno, detto di regime, in cui le grandezze in gioco hanno comportamenti detti di equilibrio, dizione questa che sarà chiarita in seguito;

(c) tanto più piccolo è Δt quanto più computazionalmente pesante è la valutazione delle equazioni descrittive del modello mentre tanto più grande è Δt quanto più "imprecisi" sono gli andamenti risultanti delle grandezze tipiche del modello.

In genere il valore di Δt è espresso come un sottomultiplo di T_∞: ad esempio se il T_∞ è fissato a 100 anni il valore di Δt può essere fissato pari a 1 (un anno), 0.5 (un semestre), 0.25 (un trimestre) e così via.
Si hanno, in genere, regole empiriche per stabilire quali siano, modello per modello, i valori ottimali di T_∞ e Δt: per la prima grandezza si parte da un valore non troppo elevato e lo si incrementa (o riduce) fino a che gli andamenti di regime non risultano evidenti, per la seconda si parte, di solito, dal valore che rappresenta l'unità di misura di T_∞ e lo si riduce fino a che gli andamenti delle variabili non cessano di mostrare variazioni apprezzabili.

(b) quantitativo se si hanno serie di dati storici con cui confrontare i valori assunti dalle variabili del modello.

6. La fase successiva consiste nell'eseguire delle simulazioni variando i valori di quelle variabili che rappresentano l'influenza del mondo esterno (vedi la Nota 1.2) e che si pensano manipolabili, in modo da valutare il comportamento del modello a seguito dell'adozione di politiche diverse[5] che potrebbero essere implementate sul modello così strutturato.

7. L'ultima fase è concettualmente collegata alla fase di *valutazione* e rappresenta la fase di riflessione sui risultati di un certo numero di simulazioni su un certo numero di politiche ritenute significative cui possono seguire modifiche:

   (a) a livello di definizione del problema;
   (b) a livello di concettualizzazione del modello;
   (c) a livello di rappresentazione formale del modello;

   che innescano di nuovo il processo a vari stadi della sua evoluzione.

Come risulta chiaro da una analisi della figura 1, il processo di definizione ed utilizzo di un modello è caratterizzato da legami ciclici fra le varie fasi (o anelli di retroazione non causali, vedi la sezione 1.2) che fanno si che il processo stesso abbia un carattere iterativo mediante il quale un modello è costruito per raffinamenti successivi: data la complessità intrinseca dei problemi che vengono analizzati non è assolutamente possibile partire dalla definizione di un "modello completo" piuttosto si parte dalla definizione di un modello minimale che viene arricchito man mano in modo che il suo comportamento sia tale che il modello sia ritenuto, sulla base di dati ricavati da osservazioni della "realtà" o di considerazioni teoriche, validato. Nella figura 1 sono individuabili cinque anelli di retroazione:

1. Simulazione $\longrightarrow$ Valutazione $\longrightarrow$ Simulazione: simulazioni successive vengono effettuate variando solo i valori dei parametri T_∞ e Δt e il risultato di ogni simulazione è valutato;

2. Valutazione $\longrightarrow$ Analisi parametrica $\longrightarrow$ Valutazione: simulazioni successive vengono effettuate variando i valori delle variabili manipolabili e il risultato di ogni simulazione è valutato;

3. Valutazione $\longrightarrow$ Raffinamento $\longrightarrow$ Definizione del problema ...: la valutazione di una o più simulazioni innesca un passo di raffinamento che porta alla necessità di una migliore definizione del problema in modo da innescare un nuovo ciclo di modellazione a partire dalla fase iniziale;

[5] Con il termine **politica** indicheremo uno dei possibili insiemi di valori di tutte le variabili (manipolabili) suddette. In generale, data la relativa insensibilità di un modello a variazioni delle variabili al di sotto di una certa soglia, una politica è un insieme di intervalli, uno per ognuna delle variabili manipolabili. La insensibilità indica che un comportamento non varia, almeno da un punto di vista qualitativo, se i parametri che possono influenzarlo non variano più di una certa quantità.

4. Valutazione $\longrightarrow$ Raffinamento $\longrightarrow$ Concettualizzazione . . . : la valutazione di una o più simulazioni innesca un passo di raffinamento che porta alla necessità di una migliore concettualizzazione del problema in modo da innescare un nuovo ciclo di modellazione dalla seconda fase;

5. Valutazione $\longrightarrow$ Raffinamento $\longrightarrow$ Rappresentazione . . . : la valutazione di una o più simulazioni innesca un passo di raffinamento che porta alla necessità di una migliore rappresentazione del problema in modo da innescare un nuovo ciclo di modellazione dalla terza fase.

Nota 1.2 *Come è ovvio che sia i modelli che sono concepiti e utilizzati considerano solo una parte di un tutto in modo che sia definito un "esterno" che agisce sul modello mediante variabili che si suppone siano modificabili "a piacimento". Tali variabili rappresentano la cosiddetta "frontiera" del modello e la loro definizione rappresenta uno dei punti più difficili e controversi perchè non sempre tali variabili agiscono al di fuori dell'influenza del sistema modellato. La necessità di definire un "dentro" e un "fuori" impedisce l'adozione di una visione olistica integrale.*

1.2 Le metafore della System Dynamics

La System Dynamics fa uso, essenzialmente, di tre metafore[6] per rappresentare i sistemi, metafore fra di loro collegate anche se non sempre in modo diretto, come sarà chiarito nel seguito. Le metafore utilizzate sono le seguenti[7]:

1. **diagrammi causal loop** (DCL), detti anche diagrammi ad anelli causali;

2. **diagrammi di flusso** (DF), detti anche diagrammi livelli÷flussi;

3. **grafici temporali** delle singole variabili o **grafici di dipendenza temporale** fra variabili, indicati genericamente come GDT.

In prima approssimazione si può affermare che:

1. i DCL sono semplici da tracciare, forse persino troppo, ma sono difficili da tradurre direttamente in equazioni descrittive del modello;

2. i DF sono più complicati da definire, richiedono una caratterizzazione delle variabili in tipi ma sono più facilmente traducibili in equazioni utilizzabili in una simulazione;

3. i GDT determinano legami fra le variabili e variazioni di variabili nel tempo e sono difficili da tradurre in una delle due forme suddette sebbene

[6]Una **metafora** è una figura retorica che trasferisce un significato proprio in uno figurato. Tipiche metafore di uso in ambito informatico sono la metafora del folder come archivio di informazioni contenute in unità distinte e quella della scrivania (o desktop) come ambiente di lavoro.

[7]Nella presente sezione si fa uso di concetti che saranno introdotti formalmente e caratterizzati solo più avanti ma che ammettono spiegazioni intuitive abbastanza semplici.

la traduzione non presenti la stessa difficoltà nei due casi dato che, noti gli andamenti delle dipendenze fra un gruppo di variabili, è possibile ricavare un DCL fra tali variabili mentre la traduzione in un DF richiede conoscenze ulteriori che permettano la tipizzazione delle variabili;

4. il passaggio da un DCL al corrispondente DF è impossibile se non sono noti i tipi delle variabili;

5. il passaggio da un DF al corrispondente DCL è sempre possibile sebbene non sia sempre possibile ricavare dai legami fra le variabili espressi in un DF il tipo di legame (di proporzionalità diretta o inversa) fra queste, a meno di non esaminare l'equazione che ne caratterizza il legame, per poterlo riprodurre, in modi che vedremo, in un DCL.

Si nota che i GDT se noti **a priori** possono essere usati per la validazione di un modello dal momento che li si può confrontare con le analoghe relazioni derivanti dalle simulazioni del modello stesso.
Fatte queste premesse si passa ad introdurre tre aspetti fondamentali che ci consentono di definire e caratterizzare in modo più formale i DCL ovvero:

1. un approccio in termini di relazioni causa ed effetto fra grandezze di un modello;

2. la considerazione di legami di retroazione fra gli elementi costitutivi di un modello;

3. il confine fra l'insieme di elementi che costituiscono il sistema (o almeno la parte di sistema) di cui definiamo il modello e ciò che rappresenta il "mondo esterno" al sistema e, per quanto ci riguarda, al modello.

1.2.1 Elementi di sintassi

Prima di proseguire è bene elencare, con qualche breve commento, quali sono gli elementi di base con cui si costruiscono i DCL. In modo formale definiamo così sia il lessico sia la sintassi riservando le sottosezioni seguenti alla descrizione della semantica.
Gli elementi costitutivi del lessico $\mathscr{L}$ sono i seguenti:

1. i nomi delle variabili ovvero **stringhe non vuote** di caratteri di una alfabeto che comprende anche le cifre e alcuni caratteri "speciali" quali _, $+$, $-$ e simili;

2. **archi orientati** con la freccia solo ad una estremità;

3. i segni $+$ e $-$ che caratterizzano gli archi orientati;

4. i segni $\oplus$ e $\ominus$ con una indicazione del verso di percorrenza (orario o antiorario) che caratterizzano gli anelli di retroazione o **cicli**.

I nomi delle variabili dovrebbero essere scelti in modo che il loro significato sia evidente in relazione al sistema di cui si stà definendo il modello per cui, ad esempio, si preferirà un nome come "tasso_di_natalità" a "tdn" in un modello in cui si descrive la dinamica di una popolazione.
Gli archi rappresentano relazioni causali fra le variabili per cui la notazione:

$$A \longrightarrow B \tag{1}$$

individua una relazione di causa÷effetto fra la variabile A (causa) e la variabile B (effetto). Il segno accanto alla testa della freccia indica il tipo di relazione. Si hanno i casi seguenti:

1. se il segno è $+$ la relazione è di proporzionalità diretta ovvero un incremento di A causa un incremento di B e lo stesso vale per una diminuzione;

2. se il segno è $-$ la relazione è di proporzionalità inversa ovvero un incremento di A causa una diminuzione di B mentre una diminuzione di A causa un incremento di B.

Esempi sono i seguenti:

$$A \longrightarrow_{+} B \tag{2}$$

$$A \longrightarrow_{-} B \tag{3}$$

Qualora un arco non abbia associato un segno il tipo della relazione fra le due variabili non è noto oppure non è stato ancora specificato. Per poter eseguire una analisi anche solo di tipo qualitativo è necessario provvedere alla sua determinazione. Qualora non ci si riesca è necessario indagare in modo più approfondito per capire se il legame ipotizzato fra le variabili esiste veramente o no.
I simboli $\oplus$ e $\ominus$ individuano la natura di un anello di retroazione fra le variabili del modello[8]. Per assegnarli si deve:

1. individuare un anello di retroazione;

2. capirne la natura.

Per quanto riguarda il punto 1. si procede in questo modo: si parte da una variabile non esogena (queste non possono, per loro natura, essere parte di un anello di retroazione) e si segue uno degli archi delle relazioni causali che da essa si originano arrivando ad una nuova variabile esogena per la quale si ripete la selezione di un arco uscente. Se, dopo un certo numero di passi, si ritorna sulla variabile da cui siamo partiti si è individuato un anello di retroazione. Se, durante il procedimento di ricerca, si tiene nota del numero di archi di tipo "$-$" che si incontrano lungo il cammino si può facilmente determinare la natura dell'anello di retroazione e perciò il segno da assegnargli. Se, infatti, tale numero è dispari l'anello è di tipo negativo (e lo si contrassegna con il simbolo $\ominus$) mentre se tale numero è nullo o pari l'anello è di tipo positivo e lo si contrassegna con il simbolo $\oplus$. In merito alla direzione di percorrenza (senso

[8]Un **anello di retroazione** è una catena di legami causali che si richiude su sè stessa.

orario o antiorario) la si contrassegna con una freccia attorno ai simboli $\oplus$ e $\ominus$ orientata nella stessa direzione del verso di percorrenza degli archi dell'anello. Si fa notare che tale indicazione non è significativa dato che una riflessione attorno ad un asse verticale dell'anello causa una inversione del verso di percorrenza dell'anello e ma non del suo segno.

Nota 1.3 *Un anello di retroazione di tipo* ***negativo*** *è detto essere di tipo compensativo e tende a mantenere stabili i valori delle variabili controllate mentre un anello di retroazione di tipo* ***positivo*** *è detto essere di tipo autorinforzante e tende a far si che le variabili controllate si allontanino sempre più dal loro valore iniziale.*

Le regole sintattiche del nostro linguaggio sono abbastanza semplici:

1. gli archi orientati congiungono coppie di variabili distinte per cui non è possibile che una variabile sia causa diretta di sè stessa ovvero non sono previsti cappi;
2. da una variabile deve entrare o uscire almeno un arco altrimenti il suo contributo al modello è nullo e la si può rimuovere;
3. da una variabile può uscire un numero qualunque di archi purchè ciascuno sia diretto verso una variabile diversa;
4. in una variabile può convergere un numero qualunque di archi purchè ciascuno provenga da una variabile diversa.

Nel seguito vedremo alcuni esempi che esemplificano l'uso di tali regole.
Se una variabile ha solo archi uscenti la si dice **esogena** perchè non subisce l'influenza delle altre variabili del modello ma le influenza. Le variabili la cui evoluzione dipende da altre variabili del modello, anche esogene, sono dette **endogene**. Dizioni alternative sono **variabili di controllo** o manipolabili per le variabili esogene e **variabili di stato** per le variabili endogene. Vedremo, infatti, nel seguito che le variabili esogene sono quelle su cui è possibile agire e che possono essere manipolate in modo da valutare l'effetto di politiche diverse sulla evoluzione del modello.
Se una variabile ha solo archi entranti può rappresentare un legame fra una parte del modello ed un'altra, legame che evita il tracciamento di archi non necessari in modo da non appesantire troppo la leggibilità di un modello.

1.2.2 Relazioni di causa ed effetto

L'approccio causale rappresenta il punto di partenza di una analisi basata sull'approccio della System Dynamics. Lo scopo è quello di rintracciare legami causali fra le grandezze che si pensa siano significative per la descrizione di una qualche porzione del mondo reale mediante un modello[9] in modo che sia

[9]Si evitano volutamente sottigliezze epistemologiche del tipo modelli di modelli e si suppone che i modelli che costruiremo siano, con qualche eccezione, in relazione con porzioni del mondo reale misurabili secondo i procedimenti delle scienze fisiche.

possibile isolare fattori chiave e le relative relazioni causali.
Un punto delicato è il seguente. Le catena causali di per sè non sono interessanti fino a che rimangono di tipo lineare. Ad esempio se si considera la seguente catena lineare di relazioni causali:

$$A \longrightarrow_{+} B \longrightarrow_{-} C \tag{4}$$

ci dice che A influenza B che, a sua volta, influenza C e che la prima relazione è di proporzionalità diretta e la seconda è di proporzionalità inversa ma la mancanza di feedback (o retroazione) fra le variabili fa si che il modello non sia molto interessante dal momento che è in genere abbastanza agevole ricavare l'influenza diretta di A su C senza dover considerare B se non per scopi descrittivi (ovvero perchè la sua presenza rende più chiaro il modello). Le catene causali sono realmente interessanti quando contengono **anelli di retroazione** che possono dare luogo ad **andamenti controintuitivi** delle variabili in essi coinvolte.
Un altro punto delicato delle relazioni causali è il seguente. Se si ha un legame del tipo:

$$A \longrightarrow_{+} B \tag{5}$$

si stabilisce una relazione di causa (A) ed effetto (B) a patto di considerare tutte le altre possibili cause inattive ovvero a patto di considerare che A sia la sola variabile che agisce sulla variabile B. In questo caso si ha che se A cresce allora anche B cresce e se A decresce allora anche B decresce. Se però avessi una situazione del genere:

$$nascite \longrightarrow_{+} Popolazione \longrightarrow_{+} morti \tag{6}$$

(il motivo dell'uso di lettere maiuscole per i nomi di alcune delle variabili sarà chiarito nella sezione dedicata ai DF) non potrei dire che la Popolazione cresce a seguito delle nascite perchè tale affermazione è falsa se le morti sono superiori alle nascite. Potrei fare tale affermazione nel caso avessi:

$$nascite \longrightarrow_{+} Popolazione \tag{7}$$

ovvero la variabile *nascite* è l'unica causa della variabile *Popolazione*. La (7) va letta nel modo seguente: tanto maggiori sono le nascite tanto maggiore è la crescita della Popolazione mentre tanto minori sono le nascite tanto minore è la crescita della Popolazione.
In merito alla 6 si ha che la Popolazione subisce due azioni contrastanti nel senso che aumenta grazie alle nascite e decresce a causa delle morti. A causa di ciò non è immediato stabilire l'andamento di Popolazione nel tempo se non sono noti gli andamenti di tali contributi. Torneremo a breve su tale esempio introducendo anelli di retroazione.
Le relazioni causa÷effetto vanno trattate con una certa cautela quando non discendono da leggi fisiche o logiche ma da considerazioni di tipo statistico per cui il legame di causalità è valido "in media" dal momento che non si possono escludere altre cause che, interagendo, possono dare luogo a comportamenti controintuitivi. Lo stesso dicasi se si vogliono stabilire relazioni di causa÷effetto

nel caso di comportamenti individuali che invece possono essere stabilite solo per popolazioni di individui, ovvero a livello aggregato.
È possibile dare esempi di come si possano specificare e tracciare relazioni causali. Come sarà chiaro dagli esempi la determinazione di tali relazioni è in molti casi problematica e diversi tipi di relazioni causali possono essere possibili.

1.3 La necessità del feedback

Come abbiamo già detto, se da un lato è necessario impostare relazioni causali in modo da poter affrontare un problema con gli strumenti della SD dall'altro tale attività da sola non è sufficiente dal momento che le catene causali[10] possono diventare arbitrariamente lunghe oppure possono dare vita ad una rete incomprensibile di legami fra le variabili. Un modo di ottenere una chiara rappresentazione di un sistema è quello di concentrare l'attenzione sulle catene circolari di relazioni causali dette anche **anelli causali** o **anelli di retroazione** o **anelli di feedback**. In tali strutture una causa iniziale può innescare una catena di legami causa÷effetto in modo da risultare causa indiretta di sè stessa. Per chiarire questo ed altri concetti si presenta un semplice esempio tratto da [RAD+83].

Esempio 1.2 *La figura 2 illustra un certo numero di legami causali nel modello di un sistema di riscaldamento di una stanza. Se la "temperatura_ambiente" si abbassa in seguito ad un abbassamento della "temperatura_esterna" si può avere una "attività_termostato" cui corrispondono, in successione, una "attività_caldaia", una "attività_radiatore" che avrà una influenza sulla "temperatura_ambiente". A sua volta "attività_caldaia" è il punto di inizio delle catene causali indicate e altre potrebbero essere concepite.*
Di tutte le catene causali l'unica ad essere veramente interessante è quella che si chiude in un anello di retroazione ed è illustrata in dettaglio in figura 3.
come è facile capire tale anello di retroazione contiene tutti gli elementi indispensabili per descrivere la relazione esistente fra la temperatura esterna (variabile esogena), la temperatura ambiente e il sistema di riscaldamento schematizzato come composto da:

1. *sistema di controllo o termostato,*
2. *sistema di produzione di calore o caldaia,*
3. *sistema effettore o radiatore.*

La figura 3 illustra i vari archi che rappresentano i legami causali con associati i segni che individuano i tipi delle relazioni stesse. In più si ha la tipizzazione del ciclo mediante il simbolo $\ominus$ e il senso di percorrenza del ciclo (senso orario): il ciclo illustrato in figura è di tipo negativo e spiega il fatto che l'azione della caldaia tende a mantenere la temperatura ambiente ad un valore prefissato rispetto alla temperatura esterna.

[10] Una **catena causale** è una successione di relazioni causali fra variabili distinte.

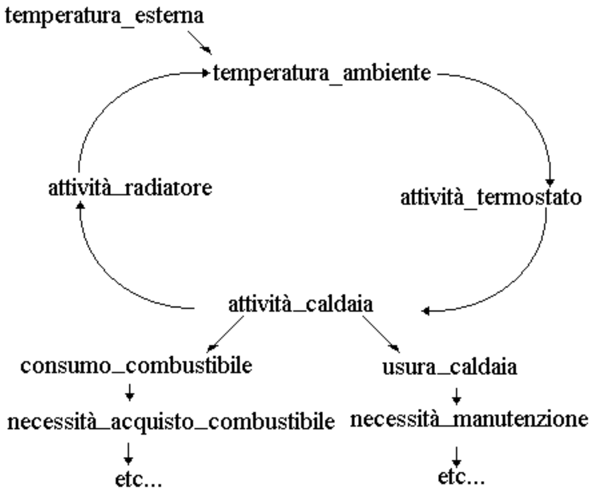

Figure 2: *Esempio di legami causali (dettagli nel testo)*

I segni sugli archi nel diagramma di figura 3 sono di facile interpretazione per cui ci si limita a dare solo qualche indicazione. Partendo dalla variabile "temperatura_esterna" si ha che tanto più questa è elevata tanto maggiore è la "temperatura_ambiente" (segno +) e di conseguenza tanto minore è "attività_termostato" (segno −) e quindi minore "attività_caldaia" (segno +) (ovvero tanto maggiore è "attività_termostato" tanto maggiore è "attività_caldaia") e così via. Con la regola vista per la tipizzazione degli anelli di retroazione, dato che il numero di archi aventi segno − è dispari l'unico anello presente è caratterizzato dal segno − e dal verso di percorrenza indicato.
L'anello di retroazione permette di controllare la temperatura ambiente ed è l'unico insieme di relazioni causali direttamente rilevante per tale controllo. Un legame causale che potrebbe avere qualche rilevanza è quello fra la disponibilità di combustibile e l'attività della caldaia dal momento che in assenza di combustibile la caldaia non può funzionare.
La ricerca di anelli di retroazione permette di individuare i confini del sistema in modo da definire il "mondo esterno" che esercita la sua azione sul modello tramite le variabili esogene. Nel caso della figura 3 il mondo esterno significativo per il sistema di riscaldamento e il modello che lo descrive è rappresentato dalla variabile esogena "temperatura_esterna".
Il fatto di limitare l'attenzione ad anelli di feedback nella costruzione di modelli per l'analisi di problemi in ambito sociale, economico o ambientale ha le seguenti conseguenze:

1. si riesce a ridurre il numero di variabili da includere in un modello ad un numero sufficientemente basso da essere gestibile mediante programmi di simulazione;

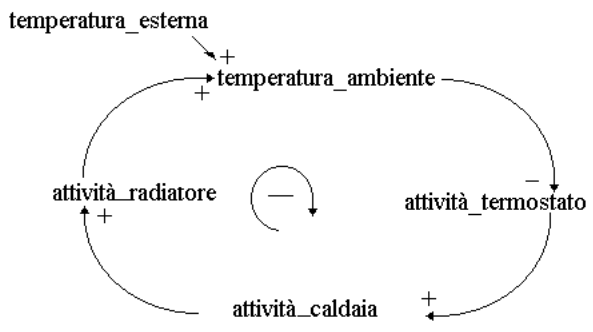

Figure 3: *Anello causale (dettagli nel testo)*

2. si può focalizzare l'attenzione solo sulle variabili che sono più importanti in relazione ai problemi che si vogliono analizzare.

Nota 1.4 *L'uso di anelli di feedback aiuta a specificare come le relazioni debbono essere rappresentate nel modello di un sistema. Un punto delicato è rappresentato dalla individuazione del confine o della frontiera del modello di un sistema. Il confine di un sistema è una sorta di linea di demarcazione che stabilisce cosa debba essere incluso nel sistema e, perciò, nel suo modello e cosa debba, invece, essere escluso (ovvero il "mondo esterno"). In più tale specificazione permette di individuare attraverso quali elementi il mondo esterno esercita la sua influenza sul sistema e sul suo modello. L'identificazione di tale confine è un processo complesso che giustifica la necessità di un processo incrementale di costruzione dei modelli e che permette di definire la dimensione, l'ambito e il carattere del problema che si stà analizzando.*
Strettamente legato a tali problematiche è il fatto che non si possono identificare gli elementi di un modello senza aver chiaro quale sia il problema che si vuole analizzare (e risolvere) e chi sia interessato al/dal problema (ovvero i cosiddetti ***stakeholder****). In molti casi, infatti, il* ***chi*** *è interessato a/da un problema definisce i confini del sistema e, quindi, del modello.*

1.4 Analisi dei *DCL*

Compito della presente sezione è quello di approfondire la conoscenza dei *DCL* come una delle metafore utilizzabili per descrivere ed analizzare la struttura di un sistema. Si tiene a sottolineare che un sistema viene identificato al termine di un processo di analisi ed astrazione che mira ad isolare da un continuum (la cosiddetta "realtà") un certo numero di fattori ritenuti essenziali per la comprensione e soluzione di un dato problema. In questa ottica la struttura di un sistema[11] viene vista come coincidente con l'insieme degli anelli di feedback

[11]E del modello corrispondente. In quanto segue ogni possibile distinzione fra un sistema e il modello che lo descrive verrà definitivamente fatta cadere. Se è vero che ad un dato sistema

che permettono di capire come mai le grandezze caratteristiche di un sistema evolvono nel tempo secondo certi andamenti.
Un modo di rappresentare la struttura di un sistema è quello di fare uso dei DCL. La rappresentazione della struttura precede, in ogni caso, la definizione delle equazioni che rappresentano i legami fra le diverse variabili del modello.

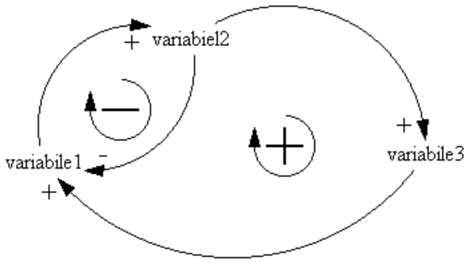

Figure 4: *Coppia di anelli causali (dettagli nel testo)*

La figura 4 presenta un generico DCL caratterizzato da[12]:

1. tre variabili ($variabile1$, $variabile2$ e $variabile3$),
2. quattro archi tipizzati (ovvero per i quali è definito il tipo di legame esistente fra le variabili presenti ai due estremi di ogni arco);
3. due cicli causali, uno *positivo* (contrassegnato con il simbolo $\oplus$, su cui è segnato, con una freccia, il verso di percorrenza degli archi del ciclo) e uno *negativo* (contrassegnato con il simbolo $\ominus$ su cui è segnato, con una freccia, il verso di percorrenza degli archi del ciclo).

L'analisi della figura 4 consente di individuare i legami fra le varie variabili e i relativi tipi nel senso che:

1. fra $variabile1$ e $variabile2$ esiste una relazione di **proporzionalità diretta**[13] ;

possono corrispondere più descrizioni ovvero più modelli, nel momento in cui un sistema viene descritto in qualche modo si stabilisce una sorta di identità fra il sistema e la sua descrizione (o modello) per cui i due termini possono essere (e saranno) considerati come sinonimi. Questa posizione nasconde implicazioni e problemi di tipo filosofico la cui trattazione esula dallo scopo delle presenti note ed è al di fuori della portata del loro autore.

[12] Tutti i diagrammi che compaiono nelle figure della presente sezione sono stati tracciati, salvo diversa specificazione, con il programma "Vensim PLE" ([Ven05]).

[13] Si ricorda che, se fra due variabili x e y esiste una relazione di proporzionalità diretta, ad un aumento/diminuzione del valore della prima corrisponde un aumento/diminuzione del valore della seconda mentre mentre, se fra due variabili x e y esiste una relazione di proporzionalità inversa, ad un aumento/diminuzione del valore della prima corrisponde una diminuzione/aumento del valore della seconda.
Una relazione del primo tipo può essere espressa matematicamente come $y = \alpha x$, con $\alpha > 0$ costante di proporzionalità, mentre una relazione del secondo tipo può essere espressa mediante una relazione matematica del tipo $y = \alpha/x$ oppure, se $x \in [0, 1]$, $y = \alpha(1 - x)$.

2. fra *variabile*2 e *variabile*1 esiste una relazione di **proporzionalità inversa**;

3. fra *variabile*2 e *variabile*3 esiste una relazione di **proporzionalità diretta**;

4. fra *variabile*3 e *variabile*1 esiste una relazione di **proporzionalità diretta**.

Nel modello di figura 4 si ha, quindi, una variabile (precisamente *variabile*1) il cui andamento nel tempo dipende da due legami causali di tipo opposto per cui non è immediato fare predizioni sul suo andamento nel tempo. Ad una prima analisi possiamo affermare, ad esempio, che:

1. se *variabile*2 aumenta[14] si ha una diminuzione di *variabile*1 e un aumento di *variabile*3;

2. all'aumento di *variabile*3 corrisponde un aumento di *variabile*1.

In tal modo l'andamento effettivo di *variabile*1 dipende da quale delle due variazioni domina sull'altra. Qualora le due variazioni si compensino a vicenda può accadere che *variabile*1 non mostri nessuna variazione nel tempo. Allo scopo di dare maggiore concretezza a quanto visto finora si passa a descrivere un semplice modello di evoluzione di una popolazione in cui il numero di individui è influenzato dai due processi concorrenti e contrastanti delle **nascite** e delle **morti**. Le *nascite* dipendono da un *tassoDiNascita* in modo che tanto più

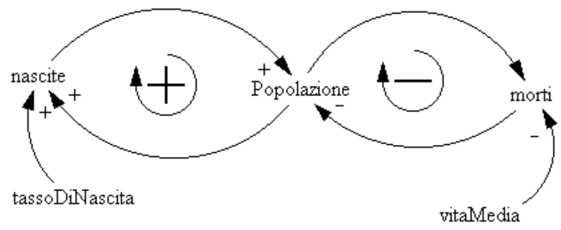

Figure 5: *DCL di una popolazione (dettagli nel testo)*

questo è alto tanto maggiore sono le prime. Le *nascite*, a loro volta, influenzano la *Popolazione* e sono da questa influenzate con relazioni di proporzionalità diretta. In questo modo si ha un ciclo causale positivo: se avessimo solo nascite la *Popolazione* crescerebbe all'infinito. Un limita alla crescita della *Popolazione* è rappresentato dal fatto che gli individui hanno una *vitaMedia* che influenza le *morti*: tanto più alta è la *vitaMedia* tanto minori sono le *morti* e viceversa. A loro volta le *morti* agiscono sulla *Popolazione* riducendola (arco con segno $-$) e *Popolazione* agisce sulle morti aumentandole (ovvero, a parità di *vitaMedia*

[14]È ovvio che ci si riferisce in questo, come in altri casi analoghi, al valore della variabile e al suo andamento nel tempo. È sperabile che queste forme concise non siano fonte di confusione.

tanto più numerosa è la *Popolazione* tanto maggiori sono le *morti* e viceversa). Si hanno così due cicli causali, uno positivo ed uno negativo, la cui interazione determina l'andamento effettivo nel tempo della *Popolazione*.
Si hanno i tre casi di figura 6:

1. l'andamento contrassegnato con 1 corrisponde al caso in cui il ciclo positivo domina su quello negativo per cui la *Popolazione* tende a crescere nel tempo a partire da un valore iniziale, anche se ad un tasso minore di quello in cui crescerebbe in assenza di mortalità;

2. l'andamento contrassegnato con 2 corrisponde al caso in cui il ciclo negativo domina su quello positivo per cui la *Popolazione* tende a decrescere nel tempo a partire da un valore iniziale;

3. l'andamento contrassegnato con 3 corrisponde al caso in cui i due cicli si bilanciano per cui la *Popolazione* rimane costante al suo valore iniziale. È facile intuire come si abbia una tale condizione di equilibrio nel caso in cui sia:
$$tasso\_di\_nascita = 1/vita\_media \tag{8}$$

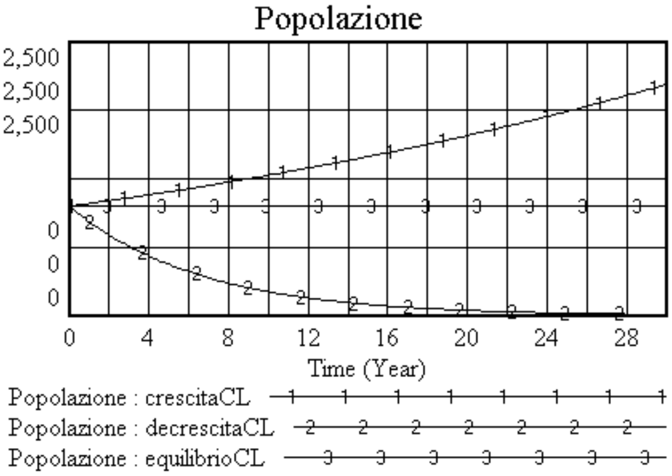

Figure 6: *Andamenti possibili di* Popolazione *(dettagli nel testo)*

Considerando che le variabili *tassoDiNascita* e *vita_media* sono variabile **esogene** (ovvero "esterne" al modello e rappresentative del "mondo esterno"), la determinazione di una condizione di equilibrio per il modello è un passo in molti casi necessario che ci consente di avere la certezza che le variazioni del comportamento del modello dipendono realmente ed esclusivamente dalle variazioni delle variabili esogene ovvero dalle "politiche" che applichiamo al modello per valutarne l'attendibilità e l'accuratezza.

La situazione descritta rappresenta una situazione ideale in cui l'andamento nel tempo della numerosità di una popolazione non ha limiti di tipo fisico rappresentati dallo spazio disponibile e dalla disponibilità di alimenti. In tal caso la dinamica dipende solo dal valore delle variabili esogene. Un primo passo[15] per rendere

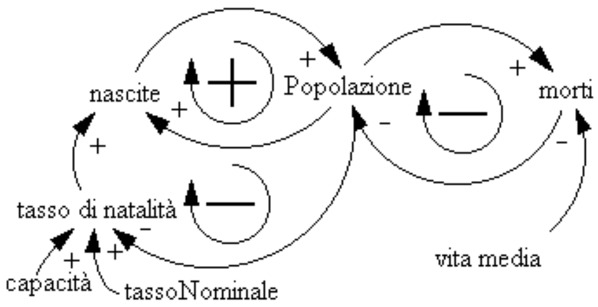

Figure 7: *Un DCL più complesso di una popolazione (dettagli nel testo)*

più realistico il modello è quello illustrato in figura 7 in cui il *tasso_di_natalità* effettivo che influenza le *nascite* dipende da:

1. un tasso di natalità teorico *tassoNominale* al quale la *Popolazione* crescerebbe in assenza di limitazioni;

2. l'entità corrente della *Popolazione*;

3. il valore massimo della *Popolazione* che l'ambiente riesce a sostenere (*capacità*).

Il modello di figura 7 contiene tre anelli causali, due dei quali sono presenti anche in figura 5 mentre il terzo contiene i legami che traducono il fatto che, al crescere della *Popolazione* verso il valore massimo, il *tasso_di_natalità* decresce fino ad un valore costante.
Gli andamenti delle variabili significative del modello sono rappresentati in figura 8. Dagli andamenti di figura 8 si vede che:

1. inizialmente l'anello causale positivo è dominante per cui le variabili *Popolazione*, *nascite* e *morti* hanno un andamento crescente;

2. la variabile *tasso_di_natalità* mostra sin da subito un andamento decrescente;

3. con il passare del tempo la differenza fra *Popolazione* e *capacità* si riduce per cui l'anello causale negativo che influisce su *tasso_di_natalità* comincia ad avere un peso sempre maggiore fino a diventare l'elemento dominante in modo che le varie variabili si assestano sui valori costanti del nuovo stato di equilibrio dinamico.

[15]Per motivi di leggibilità gli spazi interni ai nomi delle variabili presenti nelle varie figure sono sostituiti, nel testo, con il carattere"_".

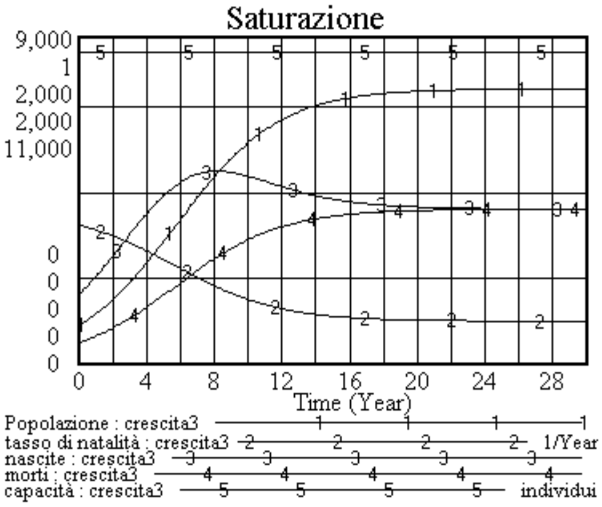

Figure 8: *Andamenti delle grandezze del modello di figura 7 (dettagli nel testo)*

Si fa notare che *Popolazione* non arriva al valore teorico massimo *capacità* a causa della presenza dei due elementi *nascite* e *morti* di pari valore (a regime) e con effetti opposti sulla variabile *Popolazione*: il primo tende ad aumentarne il valore il secondo a diminuirlo. Il modello di figura 8 è caratterizzato dalle equazioni riportate in figura 9. Una trattazione più approfondita di tali tipi di equazioni esula dagli scopi delle presenti note (al proposito si rimanda a [Ven05]). Nelle sezioni che seguono ci limiteremo ad evidenziare il significato delle equazioni chiave dei vari modelli. A questo punto della trattazione si richiama l'attenzione sulla equazione (08) che fa vedere come il *tasso_di_natalità* diminuisca al crescere di *Popolazione* mentre le equazioni (04) e (05) caratterizzano gli elementi che determinano l'andamento di *Popolazione* e l'equazione (06) fa vedere come *Popolazione* sia data da un accumulo (o integrale $INTEG$) della differenza fra *nascite* e *morti*. Un'attenta analisi delle equazioni di figura 9 permette di individuare gli anelli causali presenti nel modello e che determinano gli andamenti, per certi versi controintuitivi, mostrati in figura 8. Un modo più semplice per individuare tali anelli è quello di utilizzare le funzionalità di $Vensim$ in modo da ottenere, selezionata la variabile voluta, gli anelli causali in cui essa compare[16]. La figura 10 contiene l'elenco degli anelli causali (o *loop*) che contengono la variabile *Popolazione* e, per ciascuno di essi, la lunghezza definita come il numero di variabili presenti nell'anello e distinte dalla variabile cui il ciclo si riferisce (in questo caso *Popolazione*). Dalla figura 10 si ha la conferma del fatto che la variabile *Popolazione* è inserita in tre anelli, due dei quali condividono la variabile *nascite*. Come ultimo modello della presente sezione si

[16]Per una trattazione dei comandi utilizzabili per individuare tali anelli si rimanda alla documentazione del programma "Vensim" ([Ven05]).

```
(01)    capacità=10000
        Units: individui
(02)    FINAL TIME  = 30
        Units: Year
        The final time for the simulation.
(03)    INITIAL TIME  = 0
        Units: Year
        The initial time for the simulation.
(04)    morti=Popolazione/vita media
        Units: individui/Year
(05)    nascite=Popolazione*tasso di natalità
        Units: individui/Year
(06)    Popolazione= INTEG (+nascite-morti,1000)
        Units: individui
(07)    SAVEPER  = TIME STEP
        Units: Year [0,?]
        The frequency with which output is stored.
(08)    tasso di natalità=
           tassoNominale*(capacità-Popolazione)/capacità
        Units: 1/Year
(09)    tassoNominale=0.2
        Units: 1/Year
(10)    TIME STEP  = 0.125
        Units: Year [0,?]
        The time step for the simulation.
(11)    vita media=8
        Units: Year
```

Figure 9: *Le equazioni del modello di figura 7 (dettagli nel testo)*

presenta un semplice modello **preda÷predatore** tratto da [RAD+83] e il cui modello formale è stato tracciato con "Vensim".
Il sistema che si vuole descrivere è caratterizzato da due specie, *conigli* e *linci*: la prima costituisce l'unico cibo della seconda mentre la seconda rappresenta l'unica causa di morte della prima. La figura 11 mostra il modello nel suo complesso, in cui gli anelli causali delle due specie interagiscono fra di loro mentre la figura 12 mostra gli anelli causali delle due specie, con le semplificazioni dette, separati con in (a) il ciclo dei *conigli* e in (b) i due cicli (*nascite* e *morti*) delle *linci*. Partendo dalla figura 11 e analizzandola dall'alto verso il basso si ha:

1. un ciclo causale di tipo positivo di variazione del numero dei conigli in cui compare la variabile esogena *natalitàConigli*;
2. un ciclo causale di tipo negativo che lega i $conigliUccisiDalleLinci$ al $NumeroTotaleConigli$ con un arco di tipo $-$ e il $NumeroTotaleConigli$ ai $conigliUccisiDalleLinci$ con un arco di tipo $+$;
3. un ciclo causale di tipo positivo che lega, in successione, i $conigliUccisiDalleLinci$ alla *morteLinci* al $NumeroTotaleLinci$ per finire con i $conigliUccisiDalleLinci$;
4. un ciclo causale di tipo positivo che lega il $NumeroTotaleLinci$ alle *nasciteLinci* a loro volta dipendenti dalla variabile esogena *natalitàLinci*;

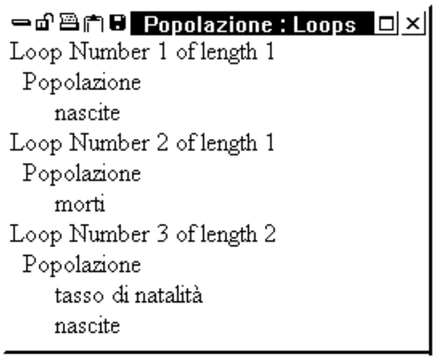

Figure 10: *Anelli causali per Popolazione (dettagli nel testo)*

5. un ciclo causale di tipo negativo che lega il *NumeroTotaleLinci* alla *morteLinci*.

I cicli presenti in figura 11 sono di facile interpretazione mentre non è altrettanto semplice predire, anche da un punto di vista qualitativo, gli andamenti delle variabili *NumeroTotaleLinci* e *NumeroTotaleConigli* nel tempo (a partire da dei valori iniziali). In questa sede ci si limita a dare una breve descrizione di due soli degli anelli causali: il secondo e il quarto (vedi l'elenco fatto in precedenza).
Il primo di tali anelli è di lunghezza pari a uno e stabilisce che tanto maggiore/minore è il numero dei *conigliUccisiDalleLinci* tanto minore/maggiore è il *NumeroTotaleConigli* mentre tanto maggiore/minore è il *NumeroTotaleConigli* tanto maggiore/minore è il numero dei *conigliUccisiDalleLinci*.
Il secondo di tali anelli è di lunghezza pari a due ed è caratterizzato da due archi di tipo $-$ che si "compensano" in modo che l'anello nel suo complesso sia di tipo positivo. Tale fatto è espressione di una regola generale: se il numero di archi di tipo $-$ contenuti in un anello causale è nullo o pari allora l'anello è di tipo positivo altrimenti è di tipo negativo.
Le semplificazioni più forti presenti nel modello sono legate al tipo di legame ipotizzato fra le due specie (i conigli sono preda solo delle linci e non hanno altre cause di mortalità e le linci si cibano solo di conigli, non hanno competitori e la loro mortalità è legata al numero di conigli che sono in grado di cacciare) e sono rese più evidenti eliminando gli anelli causali di raccordo (vedi il modello di figura 12): in tale figura si ha la parte (a) che descrive in modo irrealistico l'evoluzione delle popolazione dei conigli mentre la parte (b) descrive l'evoluzione della popolazione delle linci senza che si abbiano elementi per caratterizzare la variabile *morteLinci* (che acquista significato solo nel modello complessivo).

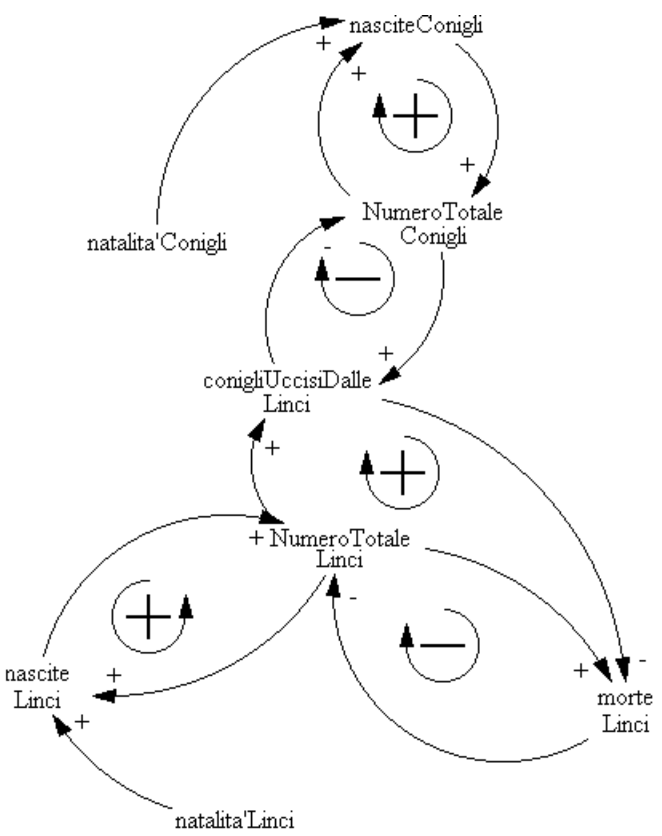

Figure 11: *Modello* ***preda÷predatore*** *completo (dettagli nel testo)*

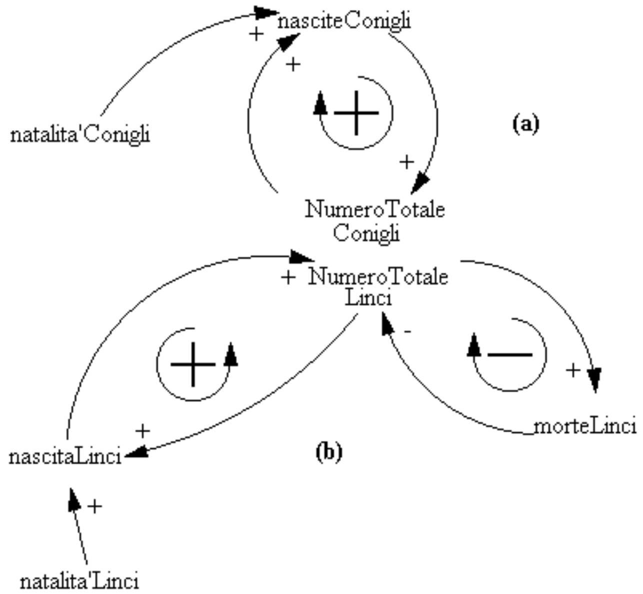

Figure 12: *Modelli **preda** e **predatore** separati (dettagli nel testo)*

1.5 Uso e significato dei *GDT*

I *GDT* possono rappresentare sostanzialmente due tipi di legami:

1. il legame fra una grandezza (che descriviamo con una **variabile**[17]) e il variare del tempo t;

2. il legame fra una grandezza e un certo numero di altre grandezze.

Gli andamenti del primo tipo sono ottenibili generalmente in due modi:

1. mediante rilevazioni "sul campo" e/o mediante l'esecuzione di test ed esperimenti con procedimenti tipici delle scienze fisiche,

2. mediante congetture di tipo teorico.

In quel caso il tempo rappresenta la "variabile indipendente" che determina le variazioni di tutte le altre variabili e può assumere valori di due tipi:

1. di tipo continuo,

2. di tipo discreto.

Nel primo caso la variabile t assume valori reali (ovvero in $\mathbb{R}$) compresi fra un istante iniziale t_0 ed un istante finale t_∞, la cui differenza individua il cosiddetto **orizzonte temporale** T_∞ mentre nel secondo caso la variabile t (denotata come t_n) assume valori multipli interi di una grandezza prefissata T (detta periodo) a partire da un istante iniziale e per un certo numero di periodi. In questo caso si ha:

$$t_n = t_0 + nT \tag{9}$$

con $n \in A \subseteq \mathbb{N}$, $A = [0, n_{max}]$. Ad esempio se si fanno rilevamenti di una popolazione con cadenza mensile per dieci anni si ha che $T = 1$, $[T] = mese$[18], $n_{max} = 120$, t_0 indica il momento di inizio dei rilevamenti e si misura il tutto in *mesi*. Se si decide, invece, di esprimere tutto in *anni* si ha $T = 1/12$, $[T] = anno$.
Gli andamenti delle variabili nel tempo rappresentano i cosiddetti **andamenti di riferimento**. Una volta definito un modello del sistema sotto esame uno dei primi passi da compiere per verificare se il modello può in qualche misura rappresentare il sistema è quello di verificare se riesce a riprodurre tali andamenti: in caso non li riproduca si può essere certi che il modello non corrisponde al sistema mentre se li riproduce non si può essere del tutto certi del fatto che il modello rappresenti il sistema che stiamo analizzando.
Gli andamenti del secondo tipo (che esprimono il legame fra una grandezza e un certo numero di altre grandezze) stabiliscono legami fra variabili, derivano in

[17] Nelle presenti note i termini **grandezza** e **variabile** saranno usati come sinonimi: una grandezza per essere descritta ha bisogno di un nome e, pertanto, ha associata una variabile mentre una variabile assumendo valori descrive una grandezza (al limite costante nel tempo). Una variabile, in genere, oltre al nome è caratterizzata da un tipo, concetto sul quale torneremo nella sezione 2.

[18] Data una grandezza x con $[x]$ se ne individua l'unità di misura, se x è adimensionale si ha $[x] = Dmnl$.

genere da congetture e teorie ma anche da misure e devono sottostare a criteri di correttezza dimensionale dal momento che le varie grandezze hanno, ciascuna, una propria unità di misura. Trascurando gli aspetti dimensionali, sui quali torneremo brevemente nella parte iniziale della Sezione 2, a questo livello ci preme osservare che tali andamenti stabiliscono relazioni fra le varie variabili in senso "aggregato" e spesso non consentono, a meno di conoscenze ulteriori, di stabilire relazioni fra una grandezza e le altre da cui dipende prese singolarmente. Il caso più semplice è quello di una variabile y che dipende da una sola variabile x tramite un fattore di conversione costante α in modo che sia:

$$y = f(\alpha, x) \tag{10}$$

In genere il legame fra le due variabili è dato da una serie di valori tabulati per cui non si dispone di una espressione matematica esplicita in forma chiusa. Conoscendo i dati della tabella si può stabilire il tipo di legame esistente fra la x e la y in modo da poterlo inserire in una struttura di tipo DCL. Qualora una variabile dipenda dall'effetto congiunto di più variabili è necessario disporre di informazioni aggiuntive per "disaccoppiare" tali effetti e stabilire per ogni variabile "causa" il tipo di legame che la lega alla variabile "effetto". In modo formale, se conosco i valori della relazione seguente:

$$y = F(x_1, x_2, x_3) \tag{11}$$

per poter definire il tipo dei legami fra la y e le x_i è necessario conoscere i valori della F in funzione delle singole x_i oltre ai valori aggregati rappresentati dalla (11).

Tornando agli andamenti delle variabili in funzione del tempo, vediamo di capire come sia possibile utilizzare tali grafici per fare inferenze sulla struttura a legami causali di un sistema ovvero di un suo modello. Come è facilmente comprensibile, perchè un tale obiettivo sia conseguibile non è sufficiente conoscere l'andamento nel tempo di una sola variabile ma è necessario conoscere almeno gli andamenti nel tempo di una coppia di variabili che si suppone siano legate da relazioni di causa÷effetto. Supponiamo ([RAD$^+$83]) di aver monitorato per un certo numero di anni l'entità di una popolazione animale (ad esempio **conigli**) nel tempo e di aver ottenuto una certa quantità di dati sotto forma di una tabella con tante entry quante sono le misurazioni disponibili, dati che sono riportati sotto forma di grafico in figura 13. I dati effettivi sono rappresentati dai puntini neri sui quali è stato sovrapposto un grafico di interpolazione. Si vuole un modello che possa giustificare in qualche modo tale andamento oscillatorio caratterizzato da un valore del periodo T e da una escursione da un valore minimo a un valore massimo abbastanza regolari.

I punti interessanti in tale grafico sono i punti in cui la popolazione di conigli raggiunge la sua massima e la sua minima grandezza e gli anni immediatamente precedenti e successivi a quelli in cui si hanno tali valori estremi. Avendo i dati relativi alla popolazione sia in forma di tabella sia in forma di grafico può essere utile calcolare le variazioni di popolazione fra due rilevazioni successive ovvero calcolare:

$$\Delta(p)_{i+1} = p(T_{i+1}) - p(T_i) \tag{12}$$

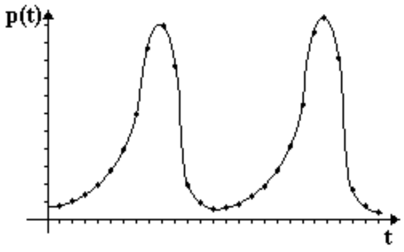

Figure 13: *Andamento nel tempo della popolazione dei* ***conigli*** *(dettagli nel testo)*

se con $p(T_i)$ si indica la popolazione all'istante T_i e con $p(T_{i+1})$ si indica la popolazione all'istante T_{i+1}, in pratica si calcola una approssimazione della derivata della funzione popolazione $p(t)$. Riportando in un grafico, in funzione del tempo discreto T_i, i valori della $\Delta(p)_{i+1}$ e interpolandoli con una curva si ottiene un grafico che incontra l'asse delle ascisse nei punti di massimo e di minimo del grafico della $p(t)$. Tale grafico ci fornisce una rappresentazione della maggiore o minore velocità di variazione della popolazione dei conigli lungo l'arco di tempo per il quale si hanno a disposizione i dati.
Fino a questo punto abbiamo descritto la modalità di variazione di una popolazione di conigli ma non il motivo perchè si ha tale variazione. In altre parole siamo in ambito descrittivo ma, poichè non abbiamo le possibili cause del fenomeno, non siamo ancora in un'ottica di analisi sistemica. Perchè ciò sia possibile si devono individuare una o più cause cui associare delle variabili e si devono stabilire delle relazioni causali con la variabile che descrive l'andamento della popolazione dei conigli nel tempo $p(t)$. Si ricorda che di tale variabile sono noti i valori solo in istanti discreti di tempo $p(T_i)$.
Possibili spiegazioni alle variazioni cicliche di $p(t)$ sono le seguenti:

1. disastri ambientali (quali siccità) che incidono sulle scorte di cibo dei conigli ma è improbabile che periodi di siccità si ripetano con tale regolarità;

2. epidemie che causano decimazioni cicliche della popolazione di conigli, eventualità improbabile perchè il periodo osservato è troppo lungo per essere associabile a cause epidemiche.

Una spiegazione plausibile dell'andamento oscillatorio della $p(t)$ è data dalla presenza, nella stessa area, di una popolazione di predatori quali le **linci**. Abbiamo già visto un modello dell'interazione fra le due popolazioni in figura 11. In questo contesto ci preme esaminare il passo precedente. Ipotizzando che la causa dell'andamento della $p(t)$ è una popolazione di linci ci servono dati in merito a questa per valutare l'andamento della sua numerosità nel tempo. Avendo a disposizione i dati relativi alla popolazione di linci sotto forma di una tabella li si può riportare sullo stesso riferimento cartesiano della $p(t)$ in modo da disegnare il grafico della $l(t)$ (i cui soli valori noti sono $l(T_i)$) insieme a quello della $p(t)$ (a meno di variazioni di scala: in genere le prede, i conigli, sono molto più numerose dei predatori, le linci). I due grafici sono riportati in figura 14. Anche una analisi superficiale delle due serie temporali di dati mostra una cor-

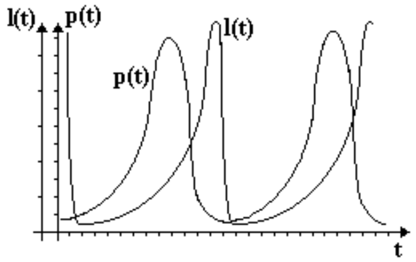

Figure 14: *Andamento nel tempo della popolazione dei* ***conigli*** *e delle* ***linci****(dettagli nel testo)*

relazione fra le due popolazioni: quando $p(t)$ è in prossimità di un minimo $l(t)$ ha un andamento decrescente mentre quando $p(t)$ cresce $l(t)$ cresce con un certo ritardo. In più, quando $l(t)$ supera una certa soglia, $p(t)$ inizia a decrescere ma $l(t)$ continua a crescere per un altro po' prima che la carenza di prede si rifletta sui predatori causando una decrescita della loro popolazione $l(t)$. Al decrescere di $l(t)$ al di sotto di una certa soglia la popolazione di conigli torna a crescere in modo che si avvii un nuovo ciclo. In questo modo si sono individuate due grandezze che descrivono l'evoluzione delle due popolazioni nel tempo e si sono iniziate a individuare delle relazioni fra di queste: le linci mangiano i conigli e il tasso di mortalità delle linci dipende dalla numerosità della popolazione di conigli. Riassumendo si ha che una eccessiva presenza di linci decima la popolazione di conigli e una carenza di conigli causa una elevata mortalità delle linci. Quando le linci sono relativamente poche i conigli si possono riprodurre con minori problemi e il loro numero tende a crescere in modo che il cibo a disposizione delle linci aumenta e con questo il numero delle linci stesse.
Dall'esempio fatto si può vedere come, partendo da delle serie temporali di dati, si possono inferire relazioni causali fra grandezze e da queste un DCL di un modello. Il procedimento necessita di cautele perchè non tutte le grandezze che entrano nel DCL (vedi la figura 11) hanno associate serie temporali di dati. Il punto chiave è il seguente: una volta creato il modello mediante la simulazione si valuta se il modello, con quella struttura, riesce a generare le serie di dati osservate nel mondo reale. In caso negativo il problema stà nel modello ovvero nelle variabili che vi abbiamo inserito e nelle relazioni che abbiamo ipotizzato. In caso affermativo non è detto che il modello sia "corretto" e, in genere, un certo grado di scetticismo si dimostra utile. Torneremo su tali concetti nella Sezione 3.

1.6 I ritardi nei modelli

In questa sezione affrontiamo un argomento particolarmente importante per la definizione dei modelli dei sistemi dinamici ovvero i **ritardi**.
I ritardi permettono di descrivere la lentezza con cui i sistemi rispondono alle sollecitazioni del mondo esterno. Nella presente sezione esamineremo le principali caratteristiche dei ritardi con l'ausilio di alcuni esempi tratti da [RAD⁺83] ed utilizzando i DCL mentre per una descrizione dei ritardi mediante i DF si rimanda alla Sezione 2.
Sin da ora si fa notare che i ritardi sono legati alla struttura del sistema ed alle sue caratteristiche fisiche, misurate dalle cosiddette **costanti di tempo**. Una **costante di tempo**, in prima approssimazione, misura la rapidità con cui un sistema risponde alle sollecitazioni passando da una situazione di regime (in cui le grandezze caratteristiche hanno certi andamenti nel tempo) ad un'altra condizione di regime (in cui le grandezze caratteristiche hanno altri andamenti nel tempo). Un ritardo, a questo livello della trattazione, rappresenta una inerzia del sistema a passare da una condizione ad un'altra.
I ritardi possono essere presenti nei modelli anche se non sono stati introdotti esplicitamente oppure, in molti casi, è necessario introdurre dei ritardi in modo da riuscire a spiegare certi andamenti riscontrati nelle serie temporali e riuscire a definire modelli di sistemi che li giustifichino.
Un semplice esempio del primo tipo, su cui torneremo nella Sezione 3, è quello illustrato in figura 15. La figura rappresenta un semplice DCL in cui una

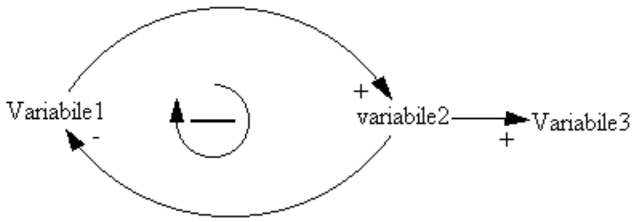

Figure 15: *Ritardi "nascosti" (dettagli nel testo)*

grandezza $Variabile1$ viene "trasferita", mediante una grandezza "intermedia" $variabile2$, ad una terza grandezza, $Variabile3$. Il diagramma non consente di capire quale sia la velocità di trasferimento e si può ipotizzare che questa sia qualunque e che il trasferimento sia, al limite, immediato. Attribuendo a $Variabile1$ un valore iniziale, ipotizzando una velocità di trasferimento finita, si capisce facilmente come tale trasferimento non possa essere immediato ma, invece, richieda del tempo. La figura 16 mostra gli andamenti delle variabili nel tempo e fa vedere che il trasferimento richieda un tempo dipendente dalla struttura del modello e dal fatto che la velocità di trasferimento non è infinita.
Esaminando la figura 16 è facile capire che la velocità iniziale di trasferimento è di $2.5unit\grave{a}/Month$. È facile intuire come la presenza di più anelli, quale quello

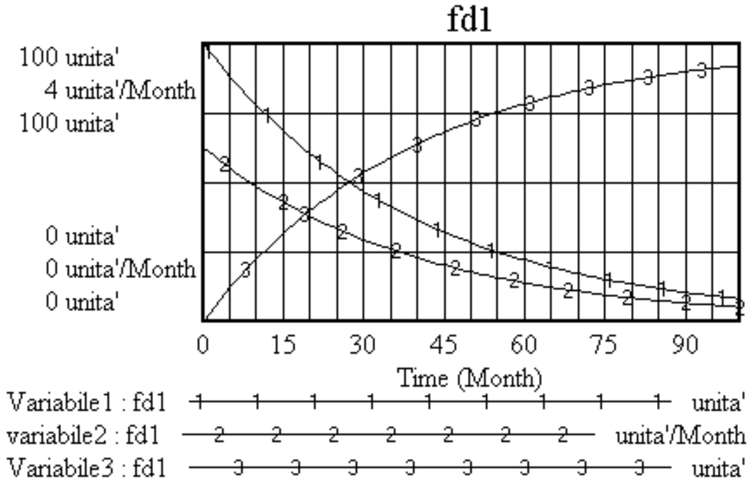

Figure 16: *Evoluzione delle grandezze del DCL di figura 15 (dettagli nel testo)*

fra $Variabile1$ e $variabile2$, in cascata non possa che tradursi in un ulteriore rallentamento del trasferimento di dati dalla variabile "iniziale" (in questo caso $Variabile1$) alla variabile "finale" (in questo caso $Variabile3$).
Vediamo ora due esempi, tratti da [RAD$^+$83], in cui si vuole illustrare come:

1. i ritardi siano in grado di giustificare gli andamenti osservati;

2. gli andamenti osservati, discordanti da quelli previsti da un modello impongano l'adozione di strutture più complesse dei modelli che, come effetto collaterale, hanno quello di introdurre ritardi "strutturali" quale quello esaminato nell'esempio di figura 15.

Supponiamo di avere rilevato nelle acque di un fiume le concentrazioni di un inquinante $p(t)$ e di aver riportato i dati, in funzione del tempo, nel grafico in alto di figura 17. In [RAD$^+$83] la storia è raccontata in dettaglio, a noi interessa l'essenziale. Conosciamo il "colpevole", una Ditta che produce prodotti chimici, e conosciamo tutto ciò che serve in merito ai prodotti e ai cicli produttivi. Gli aspetti interessanti dell'andamento di figura 17 sono i seguenti:

1. incremento rapido e istantaneo;

2. decadimento rapido fino a 0;

3. andamento rigorosamente periodico.

L'analisi svolta in [RAD$^+$83] porta alla definizione del *DCL* di figura 18. L'andamento di figura 17 descrive l'evoluzione nel tempo della variabile $livello\_di\_inquinanti\_nel\_fiume$ di figura 18. Tale variabile ha un andamento

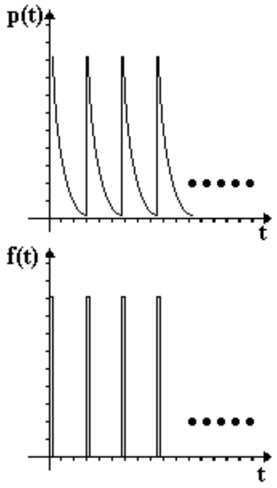

Figure 17: *Andamento della concentrazione di inquinanti (dettagli nel testo)*

governato da un ritardo simile a quello di figura 15 e dipendente da una costante di tempo che è il **tempo di dimezzamento** dell'inquinante, ovvero il tempo che una quantità di inquinante impiega per dimezzarsi. Nel caso di figura 17 tale costante di tempo ha un valore che consente l'assorbimento totale (senza permanenza di residui) dell'inquinante fra uno scarico degli impianti (vedi la figura 18) e il successivo. Qualora tale costante di tempo aumentasse (ad esempio, a seguito di variazioni della tipologia dei prodotti o dei cicli produttivi) l'inquinante non sarebbe assorbito dal fiume e comincerebbe ad accumularsi (vedi [RAD+83] per una analisi del caso). A noi interessa notare che l'anello causale negativo di figura 18 fra le variabili $livello\_d\imath_inquinanti\_nel\_fiume$ e $tasso\_di\_assorbimento\_dell'inquinante$ rappresenta un ritardo nell'andamento di decrescita del $livello\_di\_inquinanti\_nel\_fiume$ a seguito di immissioni massicce, istantanee e periodiche di inquinanti nel fiume, associate alla variabile $tasso_{di}aggiunta_{di}inquinanti_nel_fiume$ che ha l'andamento di un impulso (vedi il grafico grafico $f(t)$ in basso di figura 17). Il ritardo associato al'anello causale negativo rappresenta la lentezza con la quale la variabile $livello\_di\_inquinanti\_nel\_fiume$ reagisce all'andamento della variabile $tasso\_di\_aggiunta\_di\_inquinanti\_nel\_fiume$. In questo caso il ritardo è associato al tempo di dimezzamento dell'inquinante.
Vediamo ora un altro esempio tratto di nuovo da [RAD+83]. Di nuovo ci si concentra sull'essenziale. In questo caso si vuole far vedere come, usando una analogia male interpretata, si possa costruire un modello che non riproduce un andamento osservato a causa della sua eccessiva semplicità.
Supponiamo di piantare 10000 alberi a crescita rapida nel 1930 e di iniziare a tagliare i primi dopo 20 anni. L'andamento del tasso di taglio tt in alberi per

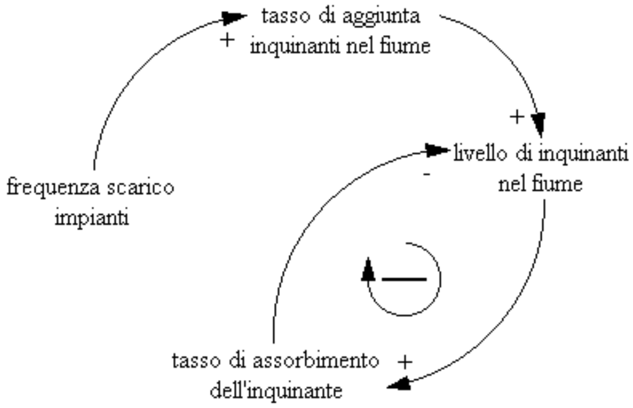

Figure 18: *DCL che "spiega" l'andamento di figura 17 (dettagli nel testo)*

anno è illustrato in figura 19. Ipotizzando un tasso di decadimento di 20 anni

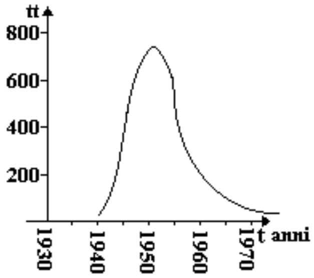

Figure 19: *Andamento del tasso_di_taglio (dettagli nel testo)*

si potrebbe essere tentati, in analogia con il caso delle figure 17 e 18, a tracciare il *DCL* di figura 20. Il problema con questo approccio è che il *DCL* è in grado di generare andamenti quali quelli di figura 17 in alto ma non quello di figura 19. Il motivo principale è che nel caso precedente si descrive una quantità accumulata nel tempo (in gergo un "livello", vedi la Sezione 2) mentre in questo caso si vuole descrivere un *tasso_di_taglio* ovvero una grandezza che si misura in *alberiperanno* (in gergo, un "flusso", vedi la Sezione 2). Tale differenza rivela che il problema stà nell'aver tracciato un *DCL* "sbagliato". La soluzione consiste nel riconoscere che gli alberi passano attraverso varie fasi di crescita e solo quelli il cui tronco ha un diametro superiore ad un valore limite sono tagliabili. Il *DCL* completo (per il quale si rimanda a [RAD$^+$83]) presenta quattro anelli causali negativi in cascata, sul primo dei quali agisce la variabile *tasso_di_piantagione*. In questo modo il *DCL* riesce a spiegare l'andamento del

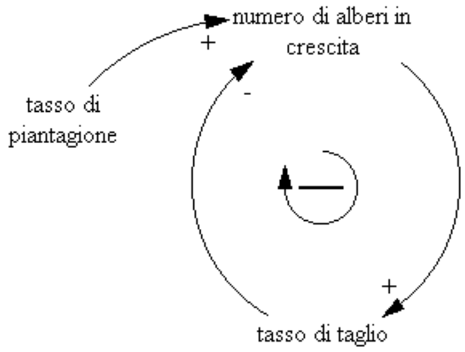

Figure 20: *Un possibile DCL per l'andamento di figura 19 (dettagli nel testo)*

tasso_di_taglio di figura 19.
Oltre ai ritardi dei tipi visti, che possiamo dire **strutturali**, esistono ritardi di tipo **esplicito** che possono essere introdotti nei modelli per svariati motivi. Torneremo su tali concetti nella Sezione 2.4.

1.7 Una parentesi filosofica: la definizione dei modelli dinamici

L'approccio della dinamica dei sistemi alla definizione di modelli ed alla loro simulazione rappresenta uno strumento per la risoluzione di problemi in quanto, dopo che un problema è stato formalizzato mediante la costruzione di un modello, consente di valutare le possibili soluzioni data una serie di obiettivi. Tutto ciò presuppone che il problema che si vuole risolvere sia stato enunciato in modo chiaro e, in particolare, che siano stati individuati gli elementi seguenti ([RAD+83]):

1. la prospettiva dalla quale si considera il problema,
2. l'orizzonte temporale lungo il quale si vuole e si può esaminare il problema,
3. quali sono gli andamenti nel tempo (ovvero gli andamenti di riferimento) significativi che abbiamo a disposizione per analizzare il problema,
4. quali politiche si possono proporre per risolvere il problema.

Ad ognuno di tali elementi corrisponde una dimensione del problema. Nel seguito, seguendo l'approccio proposto in [RAD+83], denomineremo tali dimensioni:

1. prospettiva o POV (pont of view),
2. orizzonte temporale o TF (time frame),

3. *RM* da reference mode,

4. politiche *PC* (policy choices).

Vediamo di capire il ruolo giocato dai vari elementi nel processo di costruzione di un modello.

1.7.1 *POV*

Il *POV* rappresenta il modo in cui i vari stakeholder vedono una situazione, ovvero un potenziale problema, in funzione sia del loro ruolo sia dei loro interessi. Dato un sistema i cui parametri caratteristici variano nel tempo, tali variazioni possono rappresentare un problema solo per gli stakeholder che ne sono, in qualche misura, influenzati. Un esempio fatto in [RAD+83] è quello della **crisi petrolifera** dal POV di tre stakeholder:

1. un pendolare;

2. un dirigente di una compagnia petrolifera;

3. un ambientalista.

Per il primo la ridotta disponibilità e il prezzo crescente della benzina sono un problema la cui soluzione può innescare comportamenti complessi che vanno dal car sharing alla richiesta di potenziamento del servizio pubblico su rotaia. Per il secondo la situazione rappresenta una fonte di guadagni per la sua compagnia dato che i profitti superano di gran lunga i costi di estrazione, raffinamento e trasporto. Per il terzo la crisi rappresenta un'occasione per spingere verso l'adozione di misure di risparmio energetico e diversificazione delle fonti di energia verso quelle rinnovabile e che, comunque, prescindano dall'uso del petrolio.

1.7.2 *TF*

Il *TF* indica l'arco temporale lungo il quale si ritiene che i cambiamenti siano significativi ovvero possano rappresentare un problema. In presenza di oscillazioni periodiche è necessario estendere l'analisi su più cicli mentre in presenza di fenomeni di crescita o decrescita esponenziale è necessario considerare un lasso di tempo sufficiente affinchè il fenomeno si manifesti a pieno. In genere si può fare l'analisi:

1. sul breve periodo;

2. sul lungo periodo.

Nel primo caso la presenza di oscillazioni nei cambiamenti rappresenta un elemento chiave per le decisioni che devono essere prese mentre nel secondo caso sono più significativi gli andamenti tendenziali. Nel caso visto prima della crisi petrolifera, in cui la disponibilità di prodotti petroliferi ha un andamento oscillante con una tendenza decrescente, si ha che, nel breve periodo, il problema può essere come gestire la carenza oppure quali clienti privilegiare nei periodi di disponibilità mentre sul lungo periodo entrano in gioco decisioni strategiche relative alla necessità di diversificare le fonti di approvvigionamento, di costruire depositi che agiscano da buffer fino all'uso di combustibili alternativi.

1.7.3 *RM*

I *RM* rappresentano grafici di andamenti di grandezze nel tempo relativi a:

1. osservazioni fatte e registrate;

2. comportamenti futuri ipotizzati.

Tali andamenti possono essere indicatori di problemi che si vogliono affrontare e risolvere. È necessario individuare in modo preciso quali sono gli andamenti che sono visti come indesiderabili in modo che sia possibile individuare gli elementi del modello su cui agire per apportare le volute correzioni. Si parla di **reference mode** dal momento che rappresenta un riferimento per la costruzione del modello sul quale mettere alla prova le politiche per la correzione del problema. Dato un sistema (ad esempio un lago o un fiume) relativamente al quale si vuole analizzare un problema (rispettivamente eutrofizzazione o riduzione della portata) si possono prendere in considerazione gli andamenti nel tempo di grandezze idonee a descrivere il problema ovvero:

1. nel primo caso ([RAD$^+$83]) grandezze quali la concentrazione del fosforo nell'acqua, la concentrazione di clorofilla nell'acqua, la trasparenza dell'acqua e così via;

2. nel secondo caso la piovosità nel bacino idrografico, l'entità degli emungimenti a scopi irrigui, civili e industriali, l'entità dei prelievi di inerti lungo il corso e sugli affluenti e così via.

Tali andamenti definiscono il **reference mode** del sistema e la loro definizione è un modo di formalizzare il problema. A questo punto è necessario costruire un modello del sistema in grado di riprodurre il reference mode, utilizzando gli strumenti descritti nelle presenti note, e sul quale sperimentare diverse politiche miranti alla risoluzione del problema individuato (vedi il punto che segue).

1.7.4 *PC*

Una *PC* è un insieme di azioni con le quali si alterano i valori di alcune delle grandezze che definiscono il *RM* in modo da arrivare ad una soluzione del problema. Le *PC* dipendono dal livello di comprensione sia del problema sia del sistema dal momento che tanto maggiore è tale livello tanto più efficaci sono le *PC*. Sebbene non sia necessario definire le *PC* al momento della definizione del problema il fatto di avere un certo grado di conoscenza delle possibili *PC* aiuta nella stesura sia del *DCL* sia del *DF*.

2 I diagrammi di flusso

2.1 Introduzione

Fino ad ora abbiamo posto l'accento sulla individuazione della struttura causale di un sistema in modo da individuare il confine fra il **sistema** che vogliamo analizzare e il **mondo esterno** che lo influenza e che, in certa misura, è da esso influenzato.
Nelle sezioni che seguono ci si propone di introdurre ulteriori strumenti formali che ci consentano di legare il comportamento di un sistema con la struttura di un modello in modo da spiegare il comportamento osservato. In molti casi, infatti, si dispone dei cosiddetti **andamenti di riferimento** ovvero delle registrazioni degli andamenti di grandezze nel tempo e si vuole ricavare, da tali andamenti, un modello di relazioni causali che ci consenta di capire quali siano le cause di tali variazioni.
Esempi possibili spaziano in settori molto vasti che coprono sia le attività umane sia popolazioni animali e/o vegetali. Al primo tipo appartengono ad esempio:

1. l'andamento dei consumi di energia elettrica di una famiglia o di una cittadina o di una nazione;
2. l'andamento dei consumi di combustibili liquidi e/o fossili di una famiglia o di una cittadina o di una nazione;
3. l'andamento dei consumi idrici di una cittadina;
4. l'andamento della produzione di rifiuti di una cittadina.

Al secondo tipo appartengono ad esempio:

1. l'andamento della numerosità della popolazione residente in una città;
2. l'andamento della numerosià di una popolazione di erbivori in un parco;
3. l'andamento della numerosità di due popolazioni, una di prede e l'altra di predatori.

La finestra temporale, ovvero l'arco di tempo per il quale sono disponibili i dati in funzione del tempo, può variare da qualche giorno alle decine di anni e, in modo proporzionale, può variare la cadenza con la quale sono disponibili i dati. Nel caso dell'andamento dei consumi di energia elettrica di una famiglia si possono avere andamenti giornalieri (medi) con rilevazioni (medie) con cadenza oraria. Nel caso dell'andamento della popolazione residente in una città si possono avere dati che coprono un secolo o più con cadenza annuale, quinquennale o decennale, in funzione di vincoli legislativi (che impongono l'effettuazione di censimenti periodici). In tutti questi casi si hanno serie di dati che coprono archi di tempo più o meno lunghi. Tali dati rappresentano serie temporali che ci consentono di capire qualcosa del sistema che stiamo analizzando e possono presentare regolarità o andamenti tipici che ci danno indicazioni sulla sua struttura interna.

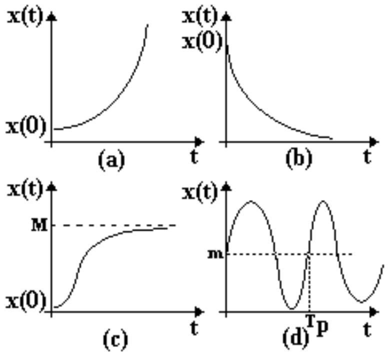

Figure 21: *Andamenti tipici (dettagli nel testo)*

La figura 21 illustra alcuni andamenti tipici e regolarità che possono essere mostrati dalle grandezza tipiche di un sistema ovvero:

1. un andamento esponenziale crescente a partire da un valore iniziale $x(0)$ (caso (a));
2. un andamento esponenziale decrescente a partire da un valore iniziale $x(0)$ (caso (b));
3. un andamento "misto" a partire da un valore iniziale $x(0)$ (caso (c));
4. un andamento oscillatorio più o meno regolare (caso (d)).

Nel caso (a) la variabile $x(t)$ ha un andamento descritto da una equazione del tipo:

$$x(t) = x(0)e^{\frac{t}{\tau}} \tag{13}$$

in cui $x(0)$ rappresenta il valore iniziale (ovvero il valore della variabile per $t = 0$) e τ rappresenta la **costante di tempo** ovvero il tempo necessario perchè la $x(t)$ passi da $x(t_1)$ a $ex(t_1)$. Si ha, infatti:

$$x(t_1) = x(0)e^{\frac{t_1}{\tau}} \tag{14}$$

e:

$$x(t_2) = x(t_1 + \tau) = x(0)e^{\frac{t_1+\tau}{\tau}} = ex(t_1) \tag{15}$$

Il caso (b) è il "duale" del precedente: in questo caso l'andamento della variabile $x(t)$ è descrivibile con una equazione del tipo:

$$x(t) = x(0)e^{-\frac{t}{\tau}} \tag{16}$$

In questo caso (vedi la figura 22) la costante di tempo τ rappresenta il tempo necessario perchè la $x(t)$ passi da $x(t_1)$ a $x(t_1)/e$. Si ha, infatti:

$$x(t_1) = x(0)e^{-\frac{t_1}{\tau}} \tag{17}$$

e:

$$x(t_2) = x(t_1 + \tau) = x(0)e^{-\frac{t_1+\tau}{\tau}} = \frac{x(t_1)}{e} \tag{18}$$

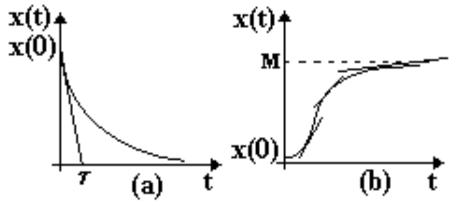

Figure 22: *Costante di tempo e variazione della pendenza (dettagli nel testo)*

Il caso caso (c) della figura 21 è detto essere "misto" perchè inizialmente la variabile mostra, a partire dal valore iniziale $x(0)$, una crescita esponenziale (e una pendenza elevata, vedi la figura 22 (b)) che si attenua con il passare del tempo (in modo che la pendenza decresca) fino a che il valore della variabile non tende ad un valore asintotico costante M. Si può dimostrare che la curva rappresentativa della $x(t)$ ha un punto di flesso in cui cambia la concavità da positiva a negativa. Tale curva viene di solito detta **logistica** e descrive fenomeni soggetti ad una qualche limitazione di tipo fisico. Nel linguaggio della Sezione 1 tale andamento è associabile a due cicli causali interagenti: uno positivo, che domina nella prima parte dell'evoluzione, e uno negativo, che domina nella seconda parte e che fa si che la crescita della variabile si assesti su un valore costante dipendente da "condizioni oggettive" e limiti fisici (spazio, cibo etc.), come sarà chiarito nel seguito anche con alcuni esempi. Il caso (d) della

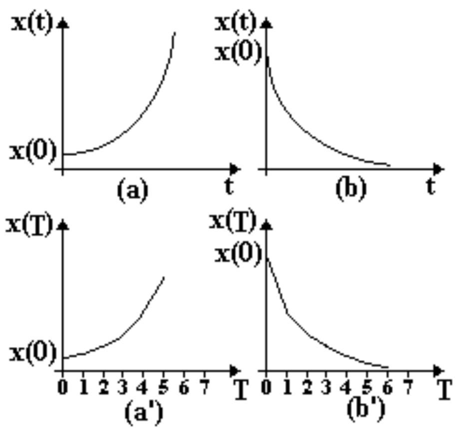

Figure 23: *Tempo continuo e tempo discreto (dettagli nel testo)*

figura 21 rappresenta un andamento oscillatorio attorno ad un valore medio m caratterizzato da una “quasi periodicità” (esaltata in figura per esigenze illustrative): la variabile assume valori (quasi) identici in istanti che differiscono per multipli di una (quasi) costante detta **periodo** e indicata con T_p. Si noti che da un punto di vista matematico il periodo è una costante ma che, nei fenomeni fisici reali, difficilmente si riscontra una tale caratteristica. Il caso caso (d) della figura 21 rappresenta un fenomeno oscillatorio **non smorzato** in cui le oscillazioni mantengono una ampiezza (distanza fra il valore massimo e il valore minimo) pressappoco costante. Se tale valore tende a decrescere nel tempo fino ad annullarsi (teoricamente all'infinito, in pratica dopo una decina di costanti di tempo) si parla di **oscillazioni smorzate** mentre se tende ad aumentare senza limite (cosa che non può accadere in un sistema fisico reale) si parla di oscillazioni esponenzialmente crescenti.
La figura 23 illustra, in modo schematico, la differenza fra due variabili descritte in funzione di una variabile continua t che assume tutti i valori reali compresi fra 0 e un valore massimo (teoricamente $+\infty$) e le stesse variabili di cui sono noti i valori solo negli istanti nT, con n che assume valori interi positivi, e T che individua la distanza fra gli istanti successivi in cui tali valori sono noti mentre l'andamento fra due istanti successivi è approssimato con dei segmenti di retta.
Quanto detto finora si basa sul presupposto che le serie temporali disponibili siano tali da non presentare lacune e siano disponibili a intervalli di tempo che consentano, rappresentando graficamente i dati, di evidenziare gli andamenti tipici (o paradigmatici) sin qui descritti. Se una di queste due ipotesi non è soddisfatta può essere problematico ritrovare nelle serie temporali andamenti paradigmatici associabili a relazioni di tipo noto (che saranno introdotto nelle sezioni che seguono) fra le variabili.

2.2 Differenziare le variabili: livelli e flussi

Nei diagrammi visti nella Sezione 1 si sono rappresentati legami causali fra le varie grandezze senza fare nessun tipo di differenziazione fra le grandezze che entrano a far parte dei singoli diagrammi. Una analisi anche superficiale di tali diagrammi mostra che alcune variabili rappresentano “accumuli” di qualcosa mentre altre rappresentano “velocità di accumulo” o “velocità di svuotamento”: le prime sono dette **variabili di tipo livello** o **livelli** mentre le seconde sono dette **variabili di tipo flusso** o **flussi**. Vedremo a breve come nei modelli sia necessario introdurre variabili di altri tipi con compiti di “raccordo” sia del modello con il suo mondo esterno sia delle varie variabili livelli e flussi fra di loro.
Per il momento un metodo pratico per capire a quale categoria appartiene una variabile è il seguente. Si immagina che il tempo si fermi e si considerano le variabili: se una variabile continua ad avere un valore non nullo è un livello altrimenti no e può essere un flusso. Vedremo, infatti, che tale condizione da sola non ci permette di affermare con assoluta certezza che la variabile sia effettivamente un flusso. Tale ragionamento si basa sul fatto che i livelli siano assimilabili allo spazio e i flussi alla velocità in equazioni di moto (per cui la

regola suddetta equivale a supporre una velocità nulla). Può accadere che i livelli siano assimilabili a velocità e i flussi ad accelerazione per cui il metodo suddetto non è applicabile *sic et simpliciter* ma ci si deve porre nella condizione di accelerazione nulla (e, quindi, velocità costante).
In generale, il punto di partenza per la costruzione di un modello è costituito dalla individuazione dei **livelli**. Dal momento che sui livelli agiscono i flussi (nel senso che le variazioni dei livelli sono causate da flussi in ingresso ed in uscita da questi) il passo successivo è dato dalla individuazione di questi ultimi.
Livelli e flussi individuano e caratterizzano i cosiddetti **scambi di materia**. Gli scambi di materia sono caratterizzati, nei modelli, dalla **legge di conservazione** sia della **quantità** sia del **tipo**. Vediamo cosa questo voglia dire. La

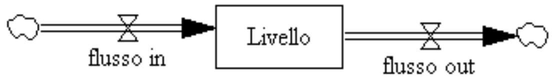

Figure 24: *Livelli, flussi e scambi di materia (dettagli nel testo)*

figura 24 mostra un semplice modello tracciato con il programma "Vensim PLE" ([Ven05])[19]. La variabile all'interno del rettangolo è un livello, le altre, vicine alle valvole, sono flussi. La legge di conservazione della materia stabilisce che[20]:

$$Livello = (flusso\_in - flusso\_out)\Delta t \tag{19}$$

in cui $\Delta t = T$ rappresenta la cadenza con cui si valutano le equazioni rappresentative del modello (vedi oltre)[21].
La legge di conservazione del tipo stabilisce che le unità di misura debbano essere compatibili per cui deve essere (se con $[x]$ si indica l'unità di misura con cui si esprime la variabile x):

$$[Livello] = [flusso\_in][T] = [flusso\_out][T] \tag{20}$$

Se ad esempio il *Livello* si misura in metri (m) e il tempo T si misura in secondi (sec), i flussi in ingresso e in uscita devono essere misurati (perchè l'equazione corrispondente sia dimensionalmente corretta) in metri al secondo (m/sec). Ogni altra combinazione è dimensionalmente scorretta o richiede l'uso di fattori di conversione per essere corretta. Se ad esempio si misura il *Livello* in chilometri (km) si deve usare il fattore di conversione:

$$1000\frac{m}{km} \tag{21}$$

[19]Tutti i diagrammi, siano essi DF o DCL, che compaiono nelle figure della presente sezione e delle successive sono stati tracciati, salvo diversa specificazione, con tale programma.

[20]Per motivi di leggibilità gli spazi interni ai nomi delle variabili presenti nelle varie figure sono sostituiti, nel testo, con il carattere"_".

[21]Nel seguito seguiremo, sperabilmente sempre, la seguente convenzione per il nomi delle variabili: i livelli avranno associate variabili con nomi la cui prima lettera è maiuscola (e lo steso vale per tutte le parole che compongono il nome), flussi e altre variabili hanno nomi con la lettera iniziale minuscola (a meno che non inizino con un acronimo).

per passare da km a m e:

$$0.001 \frac{km}{m} \tag{22}$$

per passare da m a km.
Se i flussi sono visti come portate e misurati in:

$$\frac{m^3}{h} \tag{23}$$

dove h indica l'ora, per avere compatibilità con un *Livello* misurato in metri e un T misurato in secondi si devono introdurre due fattori di conversione. Il primo permette di passare da *sec* a h ovvero è del tipo:

$$\frac{1}{3600} \frac{h}{sec} \tag{24}$$

il secondo, nell'ipotesi semplificativa (non necessaria), che il "serbatoio" cui è associata la variabile *Livello* sia di forma regolare, richiede l'introduzione di un fattore "area" A misurato in m^2 in modo che sia:

$$[Livello][A] = [flusso\_in][T] \frac{1}{3600} \frac{h}{sec} = [flusso\_out][T] \frac{1}{3600} \frac{h}{sec} \tag{25}$$

Al fine di collegare l'approccio visto nella Sezione 1 con quello della presente si riporta nella figura 25 il DCL corrispondente al modello della figura 24 che rappresenta un esempio di DF. Una rapida analisi della figura 25 ci permette di capire che:

1. tanto maggiore/minore è il valore di $flusso\_in$ tanto maggiore/minore è il valore di *Livello*;

2. tanto maggiore/minore è il valore di $flusso\_out$ tanto minore/maggiore è il valore di *Livello*.

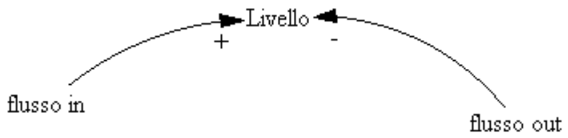

Figure 25: *Il DCL del modello di figura 24 (dettagli nel testo)*

È facile capire come sia relativamente facile passare da un DF ad un CLD (almeno nei casi più semplici) mentre il passaggio inverso non sia possibile se non si riesce a definire il tipo delle variabili coinvolte in un DCL.
Nei modelli che vedremo, pertanto, i livelli sono calcolabili conoscendo i flussi in ingresso e in uscita e valutando il loro accumulo nel tempo mentre i flussi sono calcolabili se si conosce l'andamento dei livelli nel tempo. In altre parole

i flussi non possono esistere senza i livelli (tranne che in casi patologici che noi non vedremo) e i livelli non possono esistere senza flussi se si vuole descrivere il comportamento dinamico di un sistema. Si fa notare che analizzando un sistema dinamico e il suo comportamento si ha che:

1. i livelli e i loro andamenti interessano i politici,
2. le politiche sono mirate ad influenzare i flussi,

per cui risulta di enorme importanza poter capire come livelli e flussi si influenzano a vicenda.
Oltre alle variabili livello e flusso e gli associati flussi conservativi di materia in un DF sono presenti anche altri tipi di variabili e flussi di un altro tipo. Rimandando ad una successiva sezione l'analisi degli altri tipi di variabili ci si limita ad introdurre il concetto di **flusso di informazioni**.
Un flusso di informazioni è un trasferimento **non** vincolato dalla legge di conservazione della materia fra due variabili qualunque che deve soddisfare, comunque, gli stessi principi di coerenza dimensionale cui devono sottostare i flussi di materia. Un **flusso di informazioni** o **flusso informativo** stabilisce modi immateriali mediante i quali una variabile può influenzare il comportamento di altre. La figura 26 presenta il modello di figura 24 cui sono stati aggiunti due

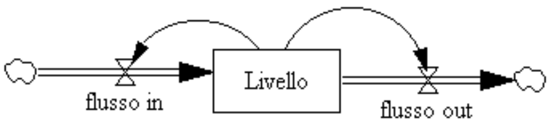

Figure 26: *Livelli, flussi, flussi di materia e di informazioni (dettagli nel testo)*

flussi di informazione che si originano entrambi dalla variabile *Livello*. Il primo di tali flussi informativi lega la variabile *Livello* a quella $flusso\_in$ mentre il secondo di tali flussi informativi lega la variabile *Livello* a quella $flusso\_out$. Se si pensano le variabili flusso come dei controlli di valvole, è come se il *Livello* avesse la capacità di aprire o chiudere tali valvole comandando le variabili ad esse associate. La relazione matematica fra *Livello* e le variabili $flusso\_in$ e $flusso\_out$ deve, ovviamente, essere scritta caso per caso e può essere sia di proporzionalità diretta sia di proporzionalità inversa oltre a poter necessitare l'uso di altre variabili. Nel caso in cui:

1. fra *Livello* e $flusso\_in$ ci sia una relazione di proporzionalità diretta,
2. fra *Livello* e $flusso\_out$ ci sia una relazione di proporzionalità diretta,

la conversione del DF di figura 26 nel corrispondente DCL dà origine al diagramma di figura 27. In tale figura si hanno due anelli causali:

1. uno positivo $flusso\_in \longrightarrow Livello \longrightarrow flusso\_in$;

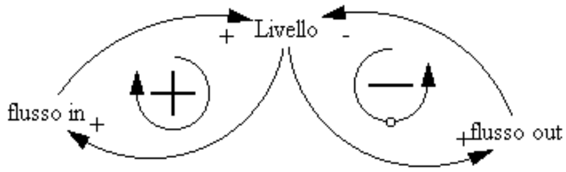

Figure 27: *Un DCL del modello di figura 26 (dettagli nel testo)*

2. uno negativo *flusso_out* $\longrightarrow$ *Livello* $\longrightarrow$ *flusso_out*;

Il comportamento effettivo di tale modello dipende dal rapporto esistente fra tali due cicli causali. Se in una prima fase domina il ciclo positivo e in una seconda il ciclo negativo si ha una crescita esponenziale della variabile *Livello* seguita da una sua decrescita. Alcuni dei comportamenti possibili sono i seguenti:

1. la variabile *Livello* non si discosta dal suo valore iniziale perchè i due flussi in ingresso e in uscita si bilanciano perfettamente (tale condizione rappresenta una condizione di equilibrio dinamico del sistema);
2. la variabile *Livello* cresce indefinitamente nel tempo;
3. la variabile *Livello* decresce indefinitamente nel tempo;
4. la variabile *Livello* decresce asintoticamente a zero.

Il secondo e il terzo comportamento (riferiti al *DCL*) non hanno plausibilità fisica e sono resi possibili dal fatto che nel modello non sono presenti gli opportuni fattori di limitazione della crescita o della decrescita. Se la variabile *Livello* rappresenta la numerosità di una popolazione animale o vegetale è evidente come questa, ad esempio, non possa crescere all'infinito (dal momento che la sua crescita sarà prima o poi ostacolata, limitata e impedita dalla carenza di spazi e di cibo) e non possa assumere valori negativi.

2.3 Differenziare le variabili: costanti e variabili ausiliarie

Utilizzando il modello che stiamo sviluppando come case study vediamo come e perchè si possano introdurre variabili di tipi diversi da quelli visti finora. Il procedimento incrementale che abbiamo seguito rappresenta un modo generale di procedere nella costruzione dei modelli. Il punto di partenza è individuare il numero minimo di livelli, poi associare ad essi i flussi di materia necessari per avere un comportamento dinamico di crescita e/o decrescita e, quindi, definire le necessarie variabili aggiuntive e i flussi informativi. L'ultimo passo è quello relativo alla stesura delle relazioni matematiche fra le varie grandezze del modello. A questo punto si procede alla risoluzione dei tali equazioni e alla determinazione degli andamenti delle varie variabili nel tempo. Una volta ottenuti tali andamenti li si confronta o con le serie temporali note o con gli andamenti attesi

e si valuta se il modello si comporta in modo coerente con le aspettative o no. In caso negativo è necessario riesaminare il modello cercando di capire cosa è stato omesso oppure cosa è stato inserito in modo improprio o inopportuno. In caso positivo non si può inferire direttamente che il modello è corretto e che la realtà, qualunque essa sia, si comporta nel modo previsto dal modello. I motivi possono essere molteplici e vanno dalla incompletezza e imprecisione dei dati sulla base dei quali si è concepito e validato il modello al fatto che il modello può essere solo un generatore di funzioni che genera le funzioni attese ma non rispecchia la struttura del sistema sotto analisi. La trattazione approfondita di tali questioni sconfina nell'epistemologia ed esula sia dallo scopo delle presenti note sia dalle competenze del loro autore. Da tali premesse discende che, ai nostri scopi, qualora un modello si comporta in modo conforme alle aspettative, siano esse di tipo fattuale (serie di dati) o teorico, lo diremo **corretto**.
A questo punto vediamo come completare il processo di costruzione del modello svolto sinora sulla base di quanto visto nella Sezione 1 e nella parte iniziale della presente. La figura 28 introduce due variabili che, nel gergo della

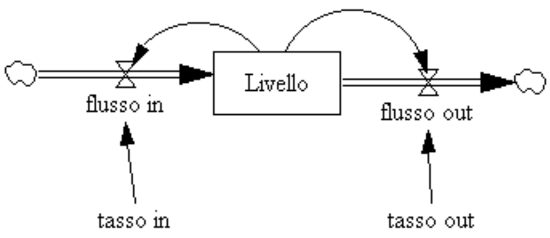

Figure 28: *Livelli, flussi, variabili "ausiliarie", flussi di materia e di informazioni (dettagli nel testo)*

System Dynamics, sono dette **costanti** ma che possono essere messe nel gruppo generico delle variabili **ausiliarie**, intendendosi con tale termine caratterizzare genericamente tutte le variabili che non sono nè livelli nè flussi. Volendo fare una tassonomia delle variabili si avrebbero i tipi seguenti:

1. livello,
2. flusso,
3. ausiliario, suddiviso in:
   (a) costante,
   (b) ausiliario,
   (c) valore iniziale,
   (d) variabile ombra.

In prima approssimazione si può affermare che:

1. le variabili di tipo **costante** permettono di rappresentare i legami fra un modello e il mondo esterno, sono le grandezze su cui si può agire per implementare e/o valutare diverse politiche sul modello che si stà costruendo;

2. le variabili di tipo **ausiliario** rappresentano i legami di tipo informativo fra i flussi i livelli e le costanti e, in molti casi, il loro scopo è principalmente quello di aumentare la leggibilità e la flessibilità di un modello;

3. le variabili di tipo **valore iniziale** permettono di manipolare direttamente i valori iniziali dei livelli in modo simile alle variabili di tipo costante così da poter valutare diverse ipotesi sulla consistenza iniziale di una grandezza modellata come un livello;

4. le variabili di tipo **variabile ombra** consentono di legare fra di loro porzioni diverse di un modello, senza che sia necessario tracciare esplicitamente un arco, in modo da aumentare la leggibilità del modello stesso. Sono utilizzabili anche per accedere esplicitamente, dall'interno delle equazioni del modello, a grandezze quali:

   (a) istante di inizio t_0 (o **INITIAL TIME**) della simulazione;

   (b) istante di fine t_∞ (o **FINAL TIME**) della simulazione;

   (c) passo di avanzamento T (o **TIME STEP**) della simulazione;

   (d) istante corrente della simulazione $time$, compreso fra t_0 e t_∞.

In figura 28 si hanno due costanti:

1. *tasso_in* che rappresenta l'influenza del mondo esterno sulla variabile $flusso\_in$,

2. *tasso_out* che rappresenta l'influenza del mondo esterno sulla variabile $flusso\_out$.

Per loro natura tali influenze sono di tipo immateriale e sono, pertanto, rappresentate da legami o flussi di tipo informativo. Oltre a svolgere una funzione di "manopola di controllo" tali variabili hanno anche la funzione di "fattori di conversione" da usarsi all'interno delle equazioni che definiscono, da un punto di vista analitico, le variabili flusso.
La figura 29 contiene un elenco di tutte le equazioni del modello di figura 28 dalle quali si può ricavare il duplice ruolo svolto dalle variabili *tasso_in* e *tasso_out* come legami con il mondo esterno (equazioni (07) e (08)) e come fattori di conversione (equazioni (02) e (03)). Esaminando la figura 29 si può vedere come sia *tasso_in* = *tasso_out*. Dalla struttura del modello si può inferire che tale eguaglianza si traduce nel fatto che la variabile *Livello* rimane fissa al suo valore iniziale 100 (vedi l'equazione (05)). La figura 30 illustra gli andamenti delle variabili principali del modello in tale condizione. Per sperimentare possibili politiche sul modello è possibile agire sulle variabili manipolabili *tasso_in* e *tasso_out*, ad esempio la prima. Se solo la variabile *tasso_in* e aumenta e passa dal valore 0.1 al valore 0.2 cosa accade al *Livello*? Si può pensare che mostri

```
(01)    FINAL TIME  = 100
        Units: Hour
        The final time for the simulation.
(02)    flusso in= Livello*tasso in
        Units: litri/Hour
(03)    flusso out= Livello*tasso out
        Units: litri/Hour
(04)    INITIAL TIME  = 0
        Units: Hour
        The initial time for the simulation.
(05)    Livello= INTEG ( flusso in-flusso out,100)
        Units: litri
(06)    SAVEPER  = TIME STEP
        Units: Hour [0,?]
        The frequency with which output is stored.
(07)    tasso in= 0.1
        Units: 1/Hour
(08)    tasso out=0.1
        Units: 1/Hour
(09)    TIME STEP  = 1
        Units: Hour [0,?]
        The time step for the simulation.
```

Figure 29: *La struttura analitica del modello (dettagli nel testo)*

una crescita, una decrescita o rimanga costante?
La figura 31 fornisce la risposta. È facile intuire che se, invece, la variabile *tasso_in* e diminuisce e passa dal valore 0.1 al valore 0.05 la variabile *Livello* mostra una decrescita. Un comportamento duale si ha se si agisce sulla variabile *tasso_out*. Se, infine, le due variabili variano nella stessa direzione e nella stessa quantità il modello rimane, ovviamente, nello stato stabile.
L'equazione (05) rappresenta la forma paradigmatica della equazione che descrive il comportamento di un livello con un flusso in ingresso e uno in uscita. In termini matematici assume la forma:

$$Livello = \int (flusso\_in - flusso\_out)\, dt + C \tag{26}$$

in cui:

1. l'operazione di integrazione $\int$ può essere vista come una somma di contributi valutati su tempi infinitesimali dt;

2. i contributi che si sommano sono la differenza fra ciò che entra ($flusso\_in$) e ciò che esce ($flusso\_out$);

3. la costante C (in questo caso pari a 100) rappresenta lo stato del livello al momento in cui si inizia la simulazione.

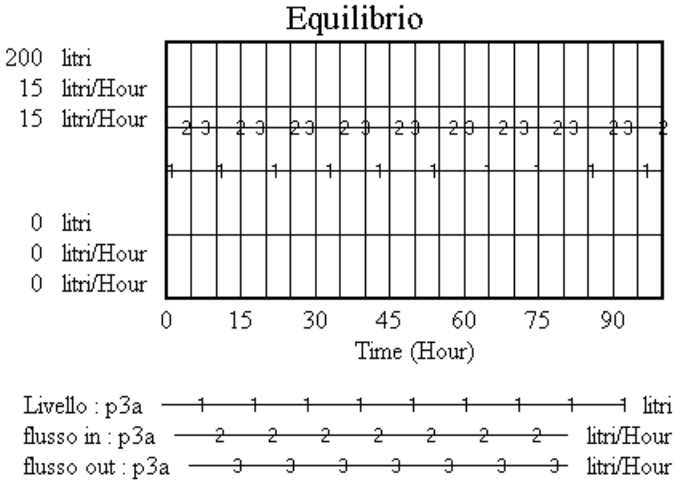

Figure 30: *Equilibrio dinamico (dettagli nel testo)*

Se *Livello* avesse solo il flusso in ingresso la sua equazione descrittiva avrebbe la forma seguente:

$$Livello = \int flusso\_in \, dt + C \tag{27}$$

ed è facile intuire come, in questo caso, si avrebbe una crescita illimitata a partire dal valore iniziale.
Se *Livello* avesse, invece, solo il flusso in uscita la sua equazione descrittiva avrebbe la forma seguente:

$$Livello = \int -flusso\_out \, dt + C \tag{28}$$

ed è facile intuire come, in questo caso, si avrebbe una decrescita illimitata a partire dal valore iniziale. Le restanti equazioni di figura 29 definiscono gli istanti di inizio e di fine della simulazione e il passo di valutazione delle varie equazioni ovvero, in pratica (ma non a rigore), il valore del *dt* che compare nelle equazioni che definiscono i livelli.
A questo punto si introducono limiti ambientali alla crescita del livello nel senso che si suppone che il *Livello* non possa crescere al di sopra di un *livello_massimo*. Per introdurre tale limitazione nel modello si fa ricorso ad un certo numero di variabili ausiliarie (vedi la figura 32) che sono:

1. $frazione\_livello$ che stabilisce il rapporto fra il valore corrente di *Livello* e il *livello_massimo*;

2. *livello_reale* che, mediante un legame funzionale permette di stabilire il valore effettivo del livello in funzione del valore corrente di $frazione\_livello$,

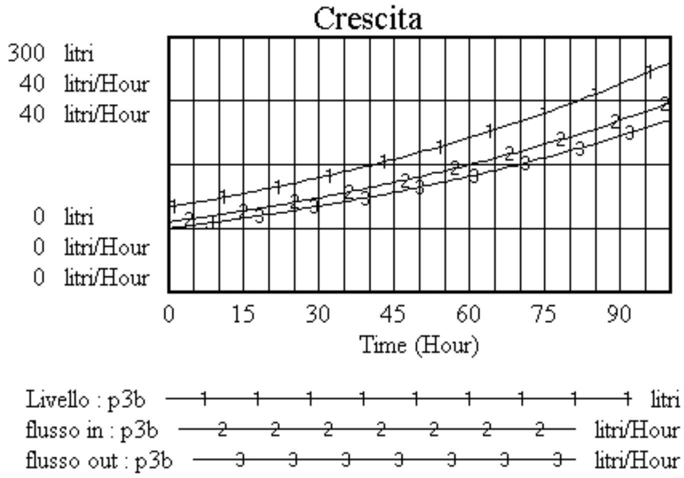

Figure 31: *Crescita (dettagli nel testo)*

3. $tasso\_effettivo$ che rapporta il $tasso\_intrinseco$ alla situazione corrente;

oltre alle seguenti costanti che rappresentano il legame con il mondo esterno:

1. $livello\_massimo$ ovvero il valore massimo cui si può arrivare;
2. $tasso\_intrinseco$ che rappresenta il tasso teorico di crescita in assenza di limitazioni ambientali;
3. $Livello0$ che rappresenta il valore iniziale della variabile $Livello$ e che si può voler variare per valutare l'effetto sul comportamento del modello di valori iniziali diversi.

Tali variabili sono gli elementi manipolabili del modello. Per studiare l'effetto di varie politiche sul $Livello$ le si può variare liberamente da una simulazione ad un'altra oppure nel corso di una simulazione. La figura 33 elenca le equazioni caratteristiche del modello appena sviluppato e rappresentato in figura 32. Una trattazione, anche solo superficiale, dei metodi per la definizione di tali equazioni mediante il programma **Vensim PLE** esula dagli ambiti delle presenti note per cui si rimanda, chi fosse interessato, alla documentazione del programma ([Ven05]). A questo livello ci preme attirare l'attenzione sulle equazioni (08) e (09).
La prima definisce $livello\_reale$ come funzione di $frazione\_livello$ ipotizzando una relazione lineare inversa (del tipo $y = 1 - x$) fra le due variabili. Tale relazione rappresenta una semplificazione che può risultare eccessiva ma che può essere un buon punto di partenza per la costruzione di un modello che include limiti ad una crescita.
La seconda definisce la variabile $Livello0$ come di tipo $INITIAL$ (ovvero un valore iniziale per un livello) e avente il valore 100.

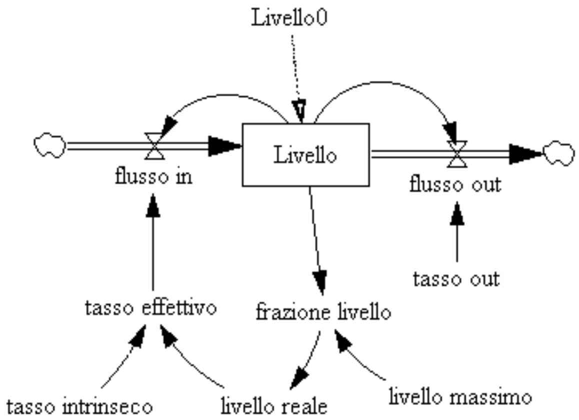

Figure 32: *Modello con limiti ambientali (dettagli nel testo)*

```
(01)  FINAL TIME  = 100
      Units: Hour
      The final time for the simulation.
(02)  flusso in= Livello*tasso effettivo
      Units: litri/Hour
(03)  flusso out= Livello*tasso out
      Units: litri/Hour
(04)  frazione livello= Livello/livello massimo
      Units: Dmnl
(05)  INITIAL TIME  = 0
      Units: Hour
      The initial time for the simulation.
(06)  Livello= INTEG ( flusso in-flusso out, Livello0)
      Units: litri
(07)  livello massimo= 1500
      Units: litri
(08)  livello reale = WITH LOOKUP ( frazione livello,([(0,0)-(1,1)],
      Units: Dmnl                                  (0,1),(1,0.0) ))
(09)  Livello0= INITIAL(100)
      Units: litri
(10)  SAVEPER  = TIME STEP
      Units: Hour [0,?]
      The frequency with which output is stored.
(11)  tasso effettivo= livello reale*tasso intrinseco
      Units: 1/Hour
(12)  tasso intrinseco= 0.5
      Units: 1/Hour
(13)  tasso out=0.1
      Units: 1/Hour [0,1,0.01]
(14)  TIME STEP  = 1
      Units: Hour [0,?]
      The time step for the simulation.
```

Figure 33: *La struttura analitica del modello (dettagli nel testo)*

Date le equazioni di figura 32, se si procede ad eseguire una simulazione, si ottengono i risultati illustrati in figura 34. In tale figura si hanno gli andamenti nel tempo (dall'istante iniziale a quello finale della simulazione) delle variabili:

1. *Livello*,
2. $flusso\_in$,
3. $flusso\_out$,
4. *livello_massimo*,

nell'ordine.

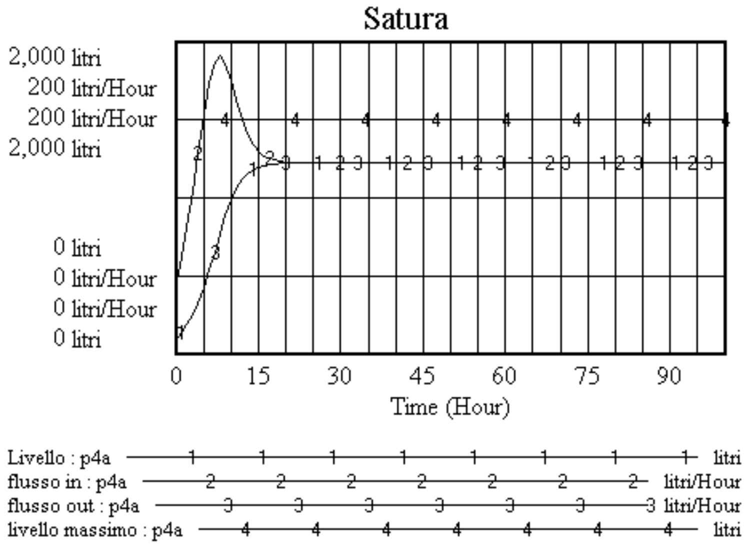

Figure 34: *Saturazione (dettagli nel testo)*

Esaminando la figura 32 si vede come, dopo un periodo di rapida crescita, la variabile *Livello* riduca il suo tasso di crescita fino a che questo non si annulla in modo che la variabile si assesti su un valore costante. Poichè inizialmente il valore di *Livello* cresce con una certa rapidità la variabile $flusso\_in$ deve crescere ad una velocità maggiore della variabile $flusso\_out$, come è mostrato in figura. Al crescere del *Livello* cresce anche la $frazione\_livello$ che agisce su *livello_reale* in modo da ridurre il $tasso\_effettivo$ e, quindi, il $flusso\_in$. Attraverso questo meccanismo, su cui torneremo, si fa in modo che limiti fisici (in questo caso il *livello_massimo*) si traducano in uno smorzamento della crescita fino a che il $flusso\_in$ (il cui tasso di crescita è sempre minore) non arriva ad eguagliare il $flusso\_out$. A quel punto i due flussi hanno pari valore e in tal modo il sistema si porta in una condizione di equilibrio dinamico. Nel caso presente, la retta che rappresenta il valore della variabile *livello_massimo* mostra come il sistema non arrivi alla capacità massima a causa della presenza, istante per istante, di un flusso in uscita non nullo. Si può verificare che, se si aumenta in modo opportuno il valore del *tasso_intrinseco*, si ottiene una crescita più ripida della variabile *Livello* e un valore di regime (o di stato stazionario) più vicino al valore *livello_massimo*, al limite coincidente con questo.

2.4 Altri esempi

Nella presente sezione si danno alcuni ulteriori esempi di sistemi dinamici per i quali si definiscono i modelli e si eseguono le simulazioni usando, come di consueto, il programma "Vensim PLE". La sezione contiene tre sottosezioni, di dimensioni diverse, dedicate, nell'ordine a:

1. esempi di uso di livelli;
2. esempi di interazioni fra flussi e livelli;
3. esempi di ritardi.

2.4.1 Livelli

I livelli rappresentano quantità che si accumulano o si smaltiscono nel tempo grazie alla interazione con i flussi in ingresso e in uscita che ne determinano le velocità di variazione. In un DF i livelli sono i primi elementi da individuare. Per ogni livello è necessario fissare:

1. l'unità di misura;
2. il valore iniziale, meglio se come variabile manipolabile.

La determinazione del valore iniziale non rappresenta, in genere, difficoltà dal momento che se si fissa una valore "inadeguato" o non si ha nessuna evoluzione oppure il modello mostra una evoluzione con un periodo transitorio più lungo ma con gli stessi andamenti di regime che si ottengono fissando valori iniziali più realistici. La determinazione dell'unità di misura richiede qualche cautela dal momento che deve essere congruente con quelle dei flussi in ingresso ed in uscita. Si ritiene degno di nota che fra l'unità di misura di un livello L e quella di un flusso f che interagisce con il livello esiste la seguente relazione:

$$[f] = \frac{[L]}{[T]} \tag{29}$$

in cui $[T]$ rappresenta l'unità di misura della variabile tempo. Da tale relazione discende immediatamente che se $[L]$ rappresenta uno "spazio" (ovvero una unità di misura che non dipende dal tempo) $[f]$ è una velocità mentre se $[L]$ rappresenta una velocità (ovvero una unità di misura che varia come l'inverso di un tempo) allora $[f]$ rappresenta una accelerazione.
I livelli sono gli elementi centrali dei modelli e sono associabili sia a flussi di materiali (che devono soddisfare vincoli di conservazione) sia a flussi di informazioni che non devono sottostare a tali vincoli. Di per sè i livelli non sono associati a comportamenti dinamici che richiedono la presenza di flussi.

2.4.2 Livelli e flussi

Vediamo ora alcune strutture semplici in cui si hanno livelli e flussi in interazione fra di loro. La figura 35 raffigura un modello con un livello e un flusso

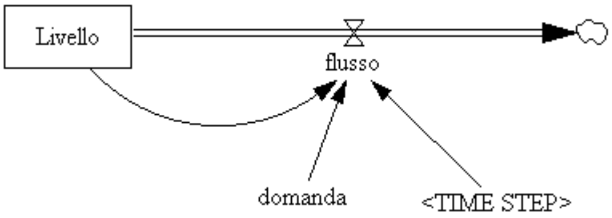

Figure 35: *Un livello e un flusso (dettagli nel testo)*

che lo "svuota". Il *Livello* parte da un valore iniziale e tende a decrescere ma non può scendere al di sotto dello 0. Il $flusso$ dipende da un parametro esterno *domanda*, dal valore corrente del *Livello* e dal valore del "quanto di tempo" $< TIME\,STEP >$. Nel modello di figura 35 si svuota *Livello* fino a che questo contiene una qauntità sufficiente a soddisfare una *domanda* costante per il tempo $< TIME\,STEP >$. Quando tale condizione non riulta più verificata il flusso va a 0 e *Livello* rimane ad un valore residuo prossimo ma non uguale a 0.
La figura 36 contiene le equazioni descrittive del modello.

(1) domanda= 2
Units: unita'/Month
(2) FINAL TIME = 100
Units: Month
The final time for the simulation.
(3) flusso= IF THEN ELSE(Livello<domanda*TIME STEP, 0 ,domanda)
Units: unita'/Month
(4) INITIAL TIME = 0
Units: Month
The initial time for the simulation.
(5) Livello= INTEG (-flusso,100)
Units: unita'
(6) SAVEPER = TIME STEP
Units: Month [0,?]
The frequency with which output is stored.
(7) TIME STEP = 1
Units: Month [0,1,0.001]
The time step for the simulation.

Figure 36: *Equazioni del modello di figura 35 (dettagli nel testo)*

L'equazione chiave è la (3) che traduce le specifiche prima evidenziate: se *Livello* non soddisfa la domanda (ovvero $Livello < domanda * TIME\,STEP$) allora il $flusso$ si annulla altrimenti il $flusso$ è pari alla *domanda*, costante e pari a 2. Gli andamenti delle variabili nel tempo sono mostrati in figura 37.

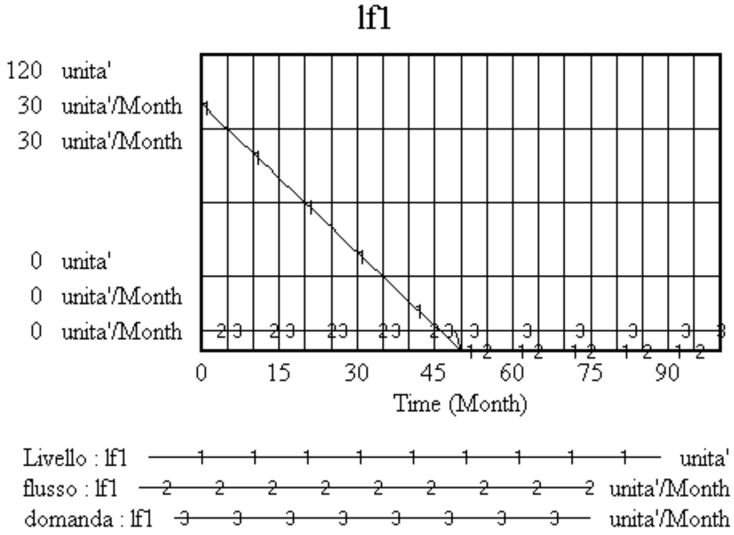

Figure 37: *Andamenti delle grandezze del modello di figura 35 (dettagli nel testo)*

Gli andamenti di figura 37 fanno vedere come la *domanda* rimanga insoddisfatta dato che *Livello* assume valori che non riescono a soddisfarla. Volendo legare la *domanda* al valore corrente del *Livello* è necessario modificare il modello come mostrato in figura 38 in cui si introducono:

1. il valore iniziale di *Livello* ovvero *Livello*0 come variabile manipolabile nelle simulazioni;

2. una variabile *domanda_nominale* che rappresenta la domanda in presenza di una risorsa illimitata;

3. una variabile *tasso_residuale* che rappresenta il complemento del tasso residuo del *Livello*;

4. una variabile *domanda* che rappresenta la domanda effettiva, funzione del valore effettivo del *Livello*

5. le variabili $flusso$ e $< TIME\,STEP >$ come nel modello precedente.

Il modello di figura 38 rappresenta una situazione in cui un *Livello* viene svuotato completamente da un $flusso$ decrescente perchè funzione di una *domanda* decrescente in quanto legata al valore corrente del *Livello* che, a sua volta, tende

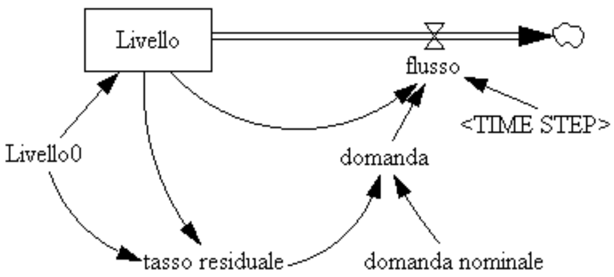

Figure 38: *Un livello e un flusso, modello modificato (dettagli nel testo)*

a decrescere. Si fa notare che nel modello di figura 35, invece, in funzione del valore di $< TIME\,STEP >$, si può avere che, nella condizione di $flusso$ nullo, *Livello* mantiene un valore residuo non nullo. Le equazioni rappresentative del modello di figura 38 hanno, come è ovvio, una complessità maggiore di quelle relative al modello di figura 35.
La figura 39 contiene un elenco delle equazioni descrittive del modello di figura 38.
Le equazioni chiave questa volta sono la (09) che definisce il $tasso\_residuale$ come:

$$tasso\_residuale = 1 - \frac{Livello0 - Livello}{Livello0} \tag{30}$$

la (01) che definisce la *domanda* come funzione del $tasso\_residuale$:

$$domanda = tasso\_residuale * domanda\_nominale \tag{31}$$

e la (04):

$$flusso = IF\,THEN\,ELSE(Livello < domanda * TIME\,STEP, 0, domanda) \tag{32}$$

```
(01)    domanda= domanda nominale*tasso residuale
        Units: unita'/Month
(02)    domanda nominale= 4
        Units: unita'
(03)    FINAL TIME  = 100
        Units: Month
        The final time for the simulation.
(04)    flusso= IF THEN ELSE(Livello<domanda*TIME STEP, 0 ,domanda )
        Units: unita'/Month
(05)    INITIAL TIME  = 0
        Units: Month
        The initial time for the simulation.
(06)    Livello= INTEG ( -flusso, Livello0)
        Units: unita'
(07)    Livello0=100
        Units: unita'
(08)    SAVEPER  =TIME STEP
        Units: Month [0,?]
        The frequency with which output is stored.
(09)    tasso residuale=1-(Livello0-Livello)/Livello0
        Units: Dmnl
(10)    TIME STEP  = 1
        Units: Month [0,1,0.001]
        The time step for the simulation.
```

Figure 39: *Equazioni del modello di figura 38 (dettagli nel testo)*

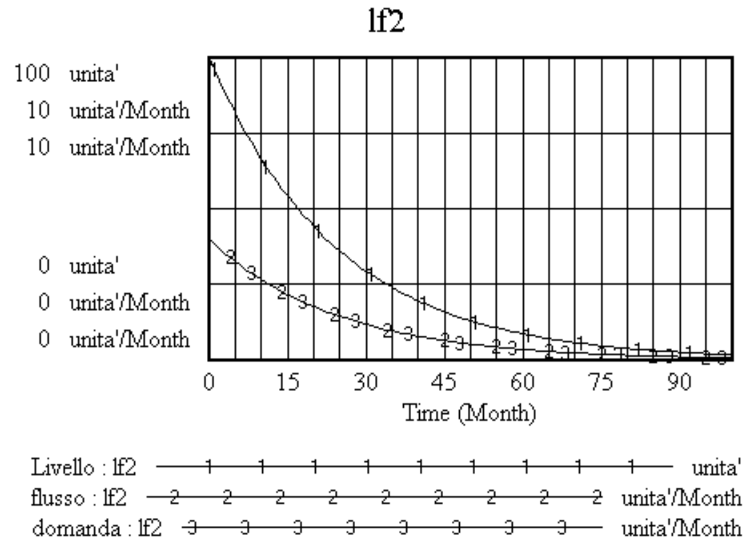

Figure 40: *Andamenti delle grandezze del modello di figura 38 (dettagli nel testo)*

2.4.3 I ritardi

I **ritardi** descrivono le inerzie, presenti nei sistemi fisici, in base alle quali:

1. non è possibile che una grandezza cambi istantaneamente e in modo discontinuo di valore;
2. se una grandezza cambia di valore una grandezza ad essa collegata da una relazione di causa÷effetto vi si adegua dopo un certo lasso di tempo.

Nei ritardi entrano in gioco le costanti di tempo τ del sistema: tanto più queste sono elevate tanto più il sistema mostra lentezza nel rispondere alle variazioni. il caso più semplice è quello di un livello con due flussi, uno di ingresso e uno di uscita. La situazione è rappresentata in figura 41. Il modello contiene un livello e due flussi. Il flusso in ingresso *in* ha un andamento impulsivo, il *Livello* (con valore iniziale $L0$) pilota il valore del flusso in uscita *out* che ripete l'andamento del flusso in ingresso con un ritardo pari a 1.

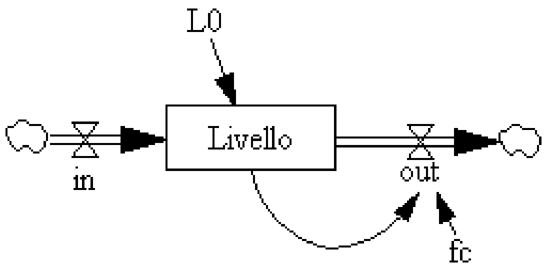

Figure 41: *Ritardo sul flusso in uscita (dettagli nel testo)*

Le equazioni descrittive di tale modello sono riportate in figura 42. Si ritengono degne di nota la (01) e la (07). La prima definisce un fattore di conversione, utile per armonizzare le unità di misura delle varie grandezze e agendo sul quale si possono variare sia l'andamento del *Livello* sia quello del flusso *out*.
La seconda definisce il valore del flusso in uscita come dipendente dal valore del *Livello* che, a sua volta, ha una espressione "canonica" rappresentata dalla (05).

(01) fc=1
Units: 1/Year
(02) FINAL TIME = 10
Units: Year
The final time for the simulation.
(03) in=50*PULSE(0 ,1)
Units: unit/Year
(04) INITIAL TIME = 0
Units: Year
The initial time for the simulation.
(05) L0=0
Units: unit
(06) Livello= INTEG (in-out,L0)
Units: unit
(07) out= Livello*fc
Units: unit/Year
(08) SAVEPER = TIME STEP
Units: Year [0,?]
The frequency with which output is stored.
(09) TIME STEP = 1
Units: Year [0,?]
The time step for the simulation.

Figure 42: *Equazioni del modello di figura 41 (dettagli nel testo)*

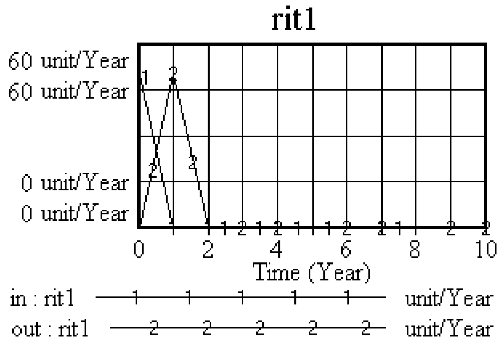

Figure 43: *Andamenti dei flussi del modello di figura 41 (dettagli nel testo)*

Un'esame degli andamenti di figura 43 lascia capire come il ritardo esistente fra il flusso in ingresso e quello in uscita sia dovuto alla presenza del *Livello* che si deve "caricare" perchè il flusso in uscita possa assumere valori non nulli. Vedremo più avanti esempi di modelli che realizzano ritardi più lunghi mediante opportune funzioni matematiche. per il momento ci si limita ad analizzare cosa accade qualora si mettano tre livelli in cascata e si applica un impulso all'ingresso del primo di tali livelli.
La situazione è quella di figura 44 in cui si hanno tre livelli in cascata, ciascuno con il suo valore iniziale, collegati fra di loro da due flussi. L'ultimo ha un flusso in uscita e il primo ha un flusso in ingresso al quale si impone un andamento impulsivo.

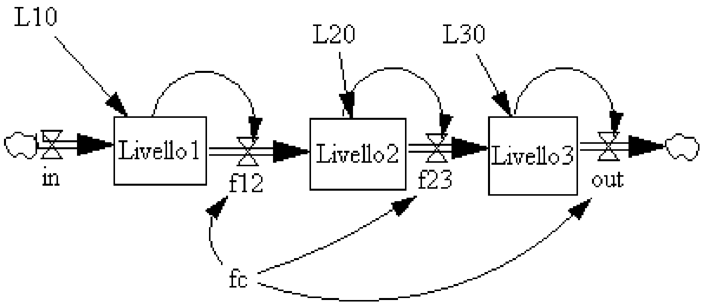

Figure 44: *Tre livelli in cascata (dettagli nel testo)*

Le equazioni descrittive del modello di figura 44 sono riportate in figura 45 mentre la figura 46 illustra il legame fra l'andamento del flusso in ingresso al primo livello e quello del flusso in uscita dall'ultimo livello. Fra i due esiste un ritardo pari al numero dei livelli intermedi. Le equazioni di figura 45 non dovrebbero presentare difficoltà interpretative.
Nel caso di figura 41 (come nel caso di figura 44) abbiamo visto cosa accade nel caso in cui sia $TIME\_STEP = 1$, pari alla durata dell'impulso di ingresso. Sarebbe interessante vedere cosa accade nei due casi seguenti:

1. $TIME\_STEP$ assume valori minori di 1 ovvero minori della durata dell'impulso in ingresso;

2. l'impulso in ingresso ha una durata maggiore di 1.

(01) f12=Livello1*fc
Units: unit/Year
(02) f23= Livello2*fc
Units: unit/Year
(03) fc=1
Units: 1/Year
(04) FINAL TIME = 10
Units: Year
The final time for the simulation.
(05) in= 50*PULSE(0 ,1)
Units: unit/Year
(06) INITIAL TIME = 0
Units: Year
The initial time for the simulation.
(07) L10= 0
Units: unit
(08) L20= 0
Units: unit
(09) L30= 0
Units: unit
(10) Livello1= INTEG (in-f12, L10)
Units: unit
(11) Livello2= INTEG (f12-f23, L20)
Units: unit
(12) Livello3= INTEG (f23-out, L30)
Units: unit
(13) out= Livello3*fc
Units: unit/Year
(14) SAVEPER = TIME STEP
Units: Year [0,?]
The frequency with which output is stored.
(15) TIME STEP = 1
Units: Year [0,?]
The time step for the simulation.

Figure 45: *Equazioni del modello di figura 44 (dettagli nel testo)*

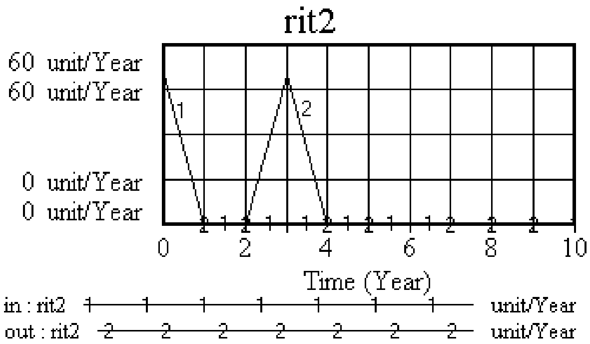

Figure 46: *Andamenti dei flussi del modello di figura 44 (dettagli nel testo)*

Il primo caso è illustrato, con riferimento al modello di figura 41, in figura 47 relativa al valore $TIME\_STEP = 0.125$. Considerazioni analoghe potrebbero essere fatte per il modello di figura 44. Il fatto di aver posto

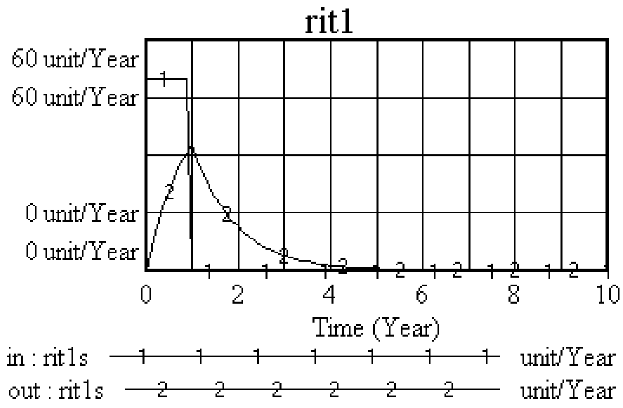

Figure 47: *Andamenti dei flussi del modello di figura 41,* $TIME\_STEP$ *ridotto (dettagli nel testo)*

$TIME\_STEP = 0.125$ con un impulso in ingresso di durata unitaria fa si che il segnale in ingresso al modello sia non nullo per 8 intervalli di tempo $TIME\_STEP$ e questo giustifica la lenta salita e il lento decadere del valore del flusso in uscita.
Più complesso è il secondo caso in cui l'impulso in ingresso ha una durata pari a $TIME\_STEP$: in questo caso, di nuovo con riferimento al modello di figura 41, si ha che la durata dell'impulso è sempre pari al passo di valutazione delle variabili del modello ovvero a $TIME\_STEP$.
Il caso di un impulso in ingesso di durata maggiore di 1 (ovvero pari a 2 (con $TIME\_STEP$ pari a 1 è illustrato in figura 48. È facile capire come, se il segnale in ingresso assume un andamento a gradino (ovvero ha un valore nullo per $t < 0$ e un valore unitario per $t \geq 0$) anche il flusso in uscita ha un andamento simile solo con un ritardo unitario.
A questo punto si introduce una categoria di ritardi realizzati mediante l'uso di funzioni matematiche ad hoc e precisamente:

1. il **ritardo pipeline**;
2. il **ritardo esponenziale**;
3. il **ritardo esponenziale del primo ordine**;
4. il **ritardo esponenziale del terzo ordine**.

Per ciascuno di tali modelli verrà esaminato il modello matematico che lo definisce e un modello realizzato con Vensim. Verranno presentate le equazioni

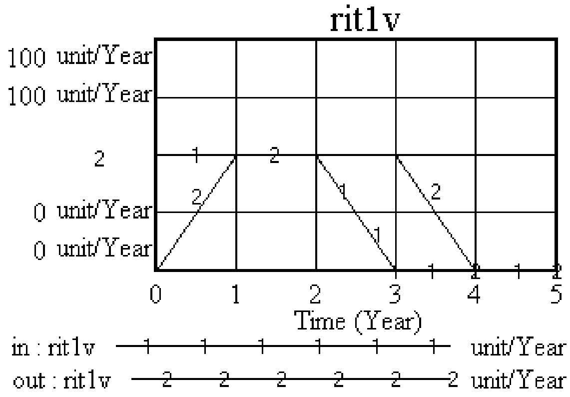

Figure 48: *Andamenti dei flussi del modello di figura 41, in di durata pari a* 2 *(dettagli nel testo)*

caratteristiche di ogni modello e almeno una simulazione, fatta utilizzando segnali (flussi) di ingresso di andamento temporale noto (i cosiddetti segnali di prova). Per ulteriori dettagli e per applicazioni di tali ritardi a problemi di tipo ambientale si rimanda a [Gal05].
Il ritardo di tipo **pipeline** si basa su una semplice relazione fra una grandezza in ingresso $in(t)$ e la corrispondente grandezza in uscita $o(t)$:

$$o(t) = i(t - \tau) \tag{33}$$

se τ rappresenta l'entità del ritardo. La figura 49 presenta un modello Vensim che realizza un ritardo di tipo **pipeline**. Le equazioni caratteristiche di tale

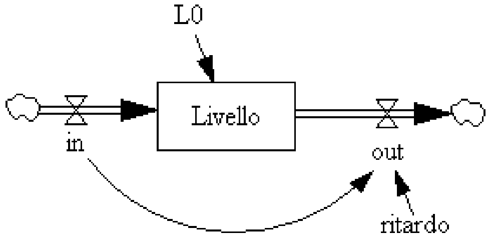

Figure 49: *Ritardo di tipo* ***pipeline*** *(dettagli nel testo)*

modello sono presentate in figura 50. L'equazione (02) definisce l'andamento del segnale in ingresso (un impulso di durata unitaria, ampiezza pari a 50 e posizionato all'istante 0), l'equazione (06) definisce l'entità del ritardo mentre la (06) definisce la relazione fra il segnale in ingresso e il segnale in uscita come determinata dalla funzione di libreria $DELAY\_FIXED$ che implementa il ritardo di tipo pipeline. Le altre equazioni hanno i significati consueti. La

```
(01)  FINAL TIME = 10
      Units: Year
      The final time for the simulation.
(02)  in=50*PULSE(0 ,1 )
      Units: unit/Year
(03)  INITIAL TIME = 0
      Units: Year
      The initial time for the simulation.
(04)  L0=0
      Units: unit
(05)  Livello= INTEG (in-out, L0)
      Units: unit
(06)  out= DELAY FIXED(in, ritardo , 0)
      Units: unit/Year
(07)  ritardo=3
      Units: Year [-1,6,1]
(08)  SAVEPER = TIME STEP
      Units: Year [0,?]
      The frequency with which output is stored.
(09)  TIME STEP = 1
      Units: Year [0,?]
      The time step for the simulation.
```

Figure 50: *Equazioni del modello di figura 49 (dettagli nel testo)*

figura 51 presenta gli andamenti nel tempo dei due segnali di ingresso e di uscita. Da tali andamenti si vede come il segnale in uscita sia una copia del segnale di ingresso traslata di una quantità pari al valore di *ritardo*. In questo modello il *Livello* ha un ruolo strumentale. Se si agisce su valore iniziale del livello, ovvero $L0$, non si ottiene nessun cambiamento dell'andamento del segnale di uscita mentre il *Livello* mostra le stesse variazioni rispetto al valore $L0$ che mostra nel caso di valore iniziale nullo. Da tutto ciò discende la differenza fra questo tipo di ritardo e quello illustrato in figura 41. Il **ritardo esponenziale** è descritto da un modello matematico più complesso. L'equazione descrittiva è una equazione differenziale del primo ordine del tipo:

$$\frac{dL(t)}{dt} = I(t) - O(t) \tag{34}$$

La (34) esprime l'equazione di bilancio di un livello $L(t)$ con un flusso in ingresso $I(t)$ e un flusso in uscita $O(t)$. Se il flusso in uscita è legato al valore corrente del livello $L(t)$ da una relazione quale la seguente:

$$O(t) = \frac{L(t)}{R} \tag{35}$$

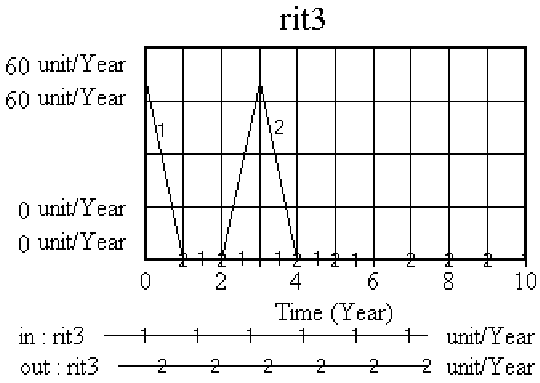

Figure 51: *Andamento dei flussi del modello di figura 49 (dettagli nel testo)*

(in cui R ha il significato di un ritardo ed assume valori strettamente positivi) è possibile scrivere la (34) sotto la seguente forma:

$$\frac{dL(t)}{dt} + \frac{L(t)}{R} = I(t) \tag{36}$$

Moltiplicando entrambi i membri della (36) per il "fattore integrale" $e^{\frac{t}{R}}$ si ottiene la seguente equazione:

$$e^{\frac{t}{R}}\frac{dL(t)}{dt} + e^{\frac{t}{R}}\frac{L(t)}{R} = e^{\frac{t}{R}}I(t) \tag{37}$$

Se si nota che a primo membro della (37) compare la derivata rispetto al tempo del termine $e^{\frac{t}{R}}L(t)$ si può scrivere la (37) come:

$$\frac{d[e^{\frac{t}{R}}L(t)]}{dt} = e^{\frac{t}{R}}I(t) \tag{38}$$

Integrando membro a membro la (38) fra 0 e t (usando s come variabile di integrazione) si ha:

$$\int_0^t \frac{d[e^{\frac{s}{R}}L(s)]}{ds} = \int_0^t e^{\frac{s}{R}}I(s)ds \tag{39}$$

ovvero:

$$e^{\frac{s}{R}}L(s) \mid_o^t = \int_0^t e^{\frac{s}{R}}I(s)ds \tag{40}$$

da cui, infine, si ottiene, con assaggi banali:

$$L(t) = e^{\frac{-t}{R}}L(0) + e^{\frac{-t}{R}}\int_0^t e^{\frac{s}{R}}I(s)ds \tag{41}$$

Per poter risolvere effettivamente la (41) è necessario conoscere una espressione in forma chiusa della $I(t)$. In queste note si considerano solo due casi semplici ([Gal05]):

1. $I(t) = 1$ per $t \in [0, +\infty)$

2. $$I(t) = \begin{cases} 1 \; t \in [0,1] \\ 0 \; t > 1 \end{cases} \tag{42}$$

Nel primo caso, in ingresso si ha una funzione a gradino unitario, che assume un valore nullo fino ad un certo istante (in questo caso $t = 0$) e un valore costante (in questo caso 1) da quell'istante in poi. Sebbene segnali di questo tipo non abbiano nessuna plausibilità fisica, li si usa spesso come segnali di prova di modelli di sistemi dinamici per testarne il comportamento. Sotto questa ipotesi la (41) assume la forma seguente:

$$L(t) = e^{\frac{-t}{R}} L(0) + e^{\frac{-t}{R}} \int_0^t e^{\frac{s}{R}} ds \tag{43}$$

ovvero:

$$L(t) = e^{\frac{-t}{R}} L(0) + e^{\frac{-t}{R}} R(e^{\frac{t}{R}} - 1) \tag{44}$$

da cui, infine:

$$L(t) = e^{\frac{-t}{R}} L(0) + R(1 - e^{\frac{-t}{R}}) \tag{45}$$

Per la (45) è facile vedere che:

1. per $t = 0$ si ha una identità;
2. per $t \to +\infty$ $L(t) = R$.

Nel secondo caso, per il quale valgono considerazioni analoghe al precedente, si devono distinguere due casi. Riprendendo la (43) qui riscritta per comodità:

$$L(t) = e^{\frac{-t}{R}} L(0) + e^{\frac{-t}{R}} \int_0^t e^{\frac{s}{R}} ds \tag{46}$$

si devono distinguere, infatti, i due casi seguenti:

1. $0 \leq t \leq 1$;
2. $t > 1$.

Nel primo caso l'integrale a secondo membro della (46) equivale al seguente:

$$\int_0^t e^{\frac{s}{R}} ds = R(e^{\frac{t}{R}} - 1) \tag{47}$$

Nel secondo caso l'integrale a secondo membro della (46) equivale al seguente:

$$\int_0^1 e^{\frac{s}{R}} ds = R(e^{\frac{1}{R}} - 1) \tag{48}$$

dato che per $t > 1$ da 1 in poi il segnale di ingresso $I(s)$ assume valore nullo. Combinando il tutto (ovvero le (47) e (48) con la (46)) si ha:

$$L(t) = e^{\frac{-t}{R}} L(0) + \begin{cases} R(e^{1-\frac{-t}{R}})\ 0 \leq t \leq 1 \\ \\ R(e^{-\frac{t-1}{R}} - e^{-\frac{t}{R}})\ t > 1 \end{cases} \tag{49}$$

Se $L(0) = 0$ la (49) si semplifica nella seguente:

$$L(t) = \begin{cases} R(e^{1-\frac{-t}{R}})\ 0 \leq t \leq 1 \\ \\ R(e^{-\frac{t-1}{R}} - e^{-\frac{t}{R}})\ t > 1 \end{cases} \tag{50}$$

Una interpretazione della (50) è la seguente:

1. per $0 \leq t \leq 1$ si ha $L'(t) > 0$ ovvero il valore del livello cresce;
2. per $t > 1$ si ha $L'(t) < 0$ e quindi il valore del livello decresce in modo che si abbia un massimo in $t = 1$.

Il valore corrente del livello cresce perchè accumula i contributi del segnale in ingresso e lo stesso vale per il valore del flusso in uscita che dipende dal valore corrente del livello. Quando, per $t > 1$, il segnale di ingresso si annulla, il livello inizia a svuotarsi e il valore del flusso in uscita a decrescere fino a che entrambi non si annullano. Il modello Vensim del ritardo esponenziale appena

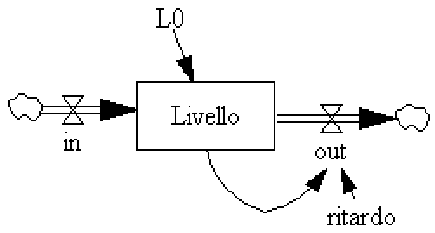

Figure 52: *Modello Vensim del ritardo esponenziale (dettagli nel testo)*

descritto è illustrato in figura 52. Le equazioni caratteristiche di tale modello sono illustrate in figura 53 mentre la figura 54 mostra gli andamenti dei flussi in ingresso e in uscita nel caso che il segnale in ingresso al modello è una funzione gradino unitario (vedi al proposito l'equazione (02) di figura 53. Se il segnale in ingresso al modello è una funzione impulso unitario gli andamenti dei flussi sono quelli riportati in figura 55. In questo caso l'equazione (02) di figura 53 è sostituita dalla seguente:

$$in = 50 * PULSE(0, 1) \tag{51}$$

```
(01)    FINAL TIME  = 10
        Units: Year
        The final time for the simulation.
(02)    in= 50*STEP(1, 0 )
        Units: unit/Year
(03)    INITIAL TIME  = 0
        Units: Year
        The initial time for the simulation.
(04)    L0= 0
        Units: unit
(05)    Livello= INTEG ( in-out, L0)
        Units: unit
(06)    out= Livello/ritardo
        Units: unit/Year
(07)    ritardo= 3
        Units: Year [-1,6,1]
(08)    SAVEPER  = TIME STEP
        Units: Year [0,?]
        The frequency with which output is stored.
(09)    TIME STEP  = 1
        Units: Year [0,?]
        The time step for the simulation.
```

Figure 53: *Equazioni del modello di figura 52 (dettagli nel testo)*

Il **ritardo esponenziale del primo ordine** può essere realizzato utilizzando una opportuna funzione matematica, fornita come primitiva da Vensim. Tale funzione ha la forma seguente:

$$output = DELAY1l(input, delay\_time, initial\_value) \qquad (52)$$

e permette di ritardare di un tempo pari a *delay_time* gli effetti di un segnale di *input* su un segnale di *output*. Il comportamento è in tutto e per tutto identico a quello del ritardo esponenziale che abbiamo esaminato in precedenza. In questo caso, la variabile *initial_value* assolve il ruolo del valore iniziale $L0$ del modello precedente.
Per completezza si riporta, in figura 56, la struttura di un modello Vensim che realizza un ritardo esponenziale del primo ordine le cui equazioni descrittive sono riportate qui di seguito:

```
(01)    FINAL TIME  = 10
        Units: Month
        The final time for the simulation.
(02)    in = PULSE(0, 1 )
        Units: unit/Month
(03)    INITIAL TIME  = 0
        Units: Month
        The initial time for the simulation.
```

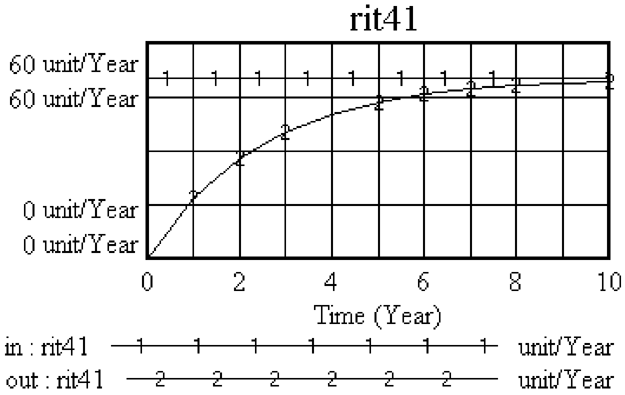

Figure 54: *Andamenti dei flussi per il modello di figura 52(dettagli nel testo)*

```
(04)    L= INTEG (in-out,L0)
        Units: unit
(05)    L0=0
        Units: unit
(06)    out=DELAY1I(in, ritardo , 0)
        Units: unit/Month
(07)    ritardo=3
        Units: Month
(08)    SAVEPER  = TIME STEP
        Units: Month [0,?]
        The frequency with which output is stored.
(09)    TIME STEP  = 1
        Units: Month [0,?]
        The time step for the simulation.
```

mentre in figura 57 è riportato l'andamento del valore del flusso in uscita in corrispondenza di un flusso in ingresso che ha l'andamento specificato dall'equazione (02). L'ultimo tipo di ritardo cui si accenna brevemente in queste note è il **ritardo esponenziale del terzo ordine**. Il **ritardo esponenziale del terzo ordine** può essere realizzato utilizzando una opportuna funzione matematica, fornita come primitiva da Vensim. Tale funzione ha la forma seguente:

$$output = DELAY3l(input, delay\_time, initial\_value) \tag{53}$$

e permette di ritardare di un tempo pari a *delay_time* gli effetti di un segnale di *input* su un segnale di *output*. Dalla documentazione ([Ven05]) si vede come un livello di tale tipo equivalga a tre ritardi di tipo esponenziale messi in cascata: ovvero tre livelli in cascata connessi da due flussi con un flusso in ingresso al primo livello e uno in uscita dall'ultimo.

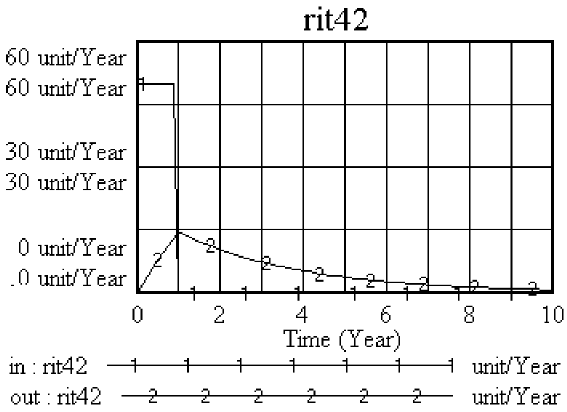

Figure 55: *Altri andamenti dei flussi per il modello di figura 52(dettagli nel testo)*

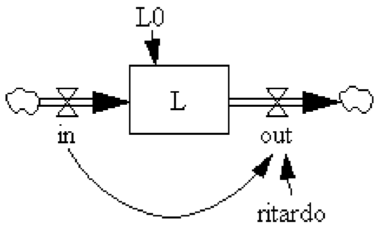

Figure 56: *Modello Vensim del ritardo esponenziale del primo ordine (dettagli nel testo)*

In figura 58, la struttura di un modello Vensim che realizza un ritardo esponenziale del terzo ordine le cui equazioni descrittive sono riportate qui di seguito:

```
(01)    FINAL TIME  = 10
        Units: Month
        The final time for the simulation.
(02)    in = PULSE(0, 1 )
        Units: unit/Month
(03)    INITIAL TIME  = 0
        Units: Month
        The initial time for the simulation.
(04)    L= INTEG (in-out,L0)
        Units: unit
(05)    L0=0
        Units: unit
(06)    out=DELAY3I(in, ritardo , 0)
```

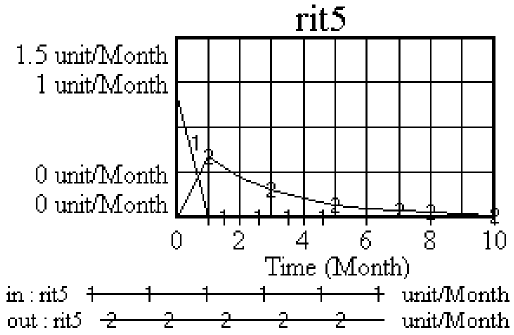

Figure 57: *Andamento dei flussi per il modello Vensim di figura 56 (dettagli nel testo)*

```
        Units: unit/Month
(07)    ritardo=3
        Units: Month
(08)    SAVEPER  = TIME STEP
        Units: Month [0,?]
        The frequency with which output is stored.
(09)    TIME STEP  = 1
        Units: Month [0,?]
        The time step for the simulation.
```

mentre in figura 59 è riportato l'andamento del valore del flusso in uscita in corrispondenza di un flusso in ingresso che ha l'andamento specificato dall'equazione (02) e con il valore di $TIME\_STEP$ specificato dall'equazione (09). Se si pone $TIME\_STEP = 0.125$ l'andamento dei flussi in ingresso e in uscita è quello illustrato in figura 60: i due diversi andamenti di *out* sono motivati dal fatto che, nel secondo caso, il segnale in ingresso assume valori non nulli per 7 "colpi di clock" (di durata ciascuno pari a $TIME\_STEP$) e si annulla solo all'ottavo, il livello e il flusso in uscita, calcolati con le opportune relazioni, assumono valori di conseguenza.

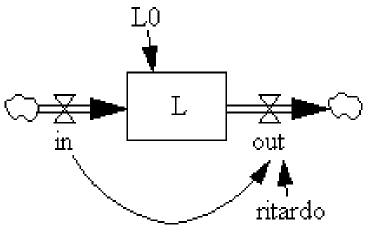

Figure 58: *Modello Vensim del ritardo esponenziale del terzo ordine (dettagli nel testo)*

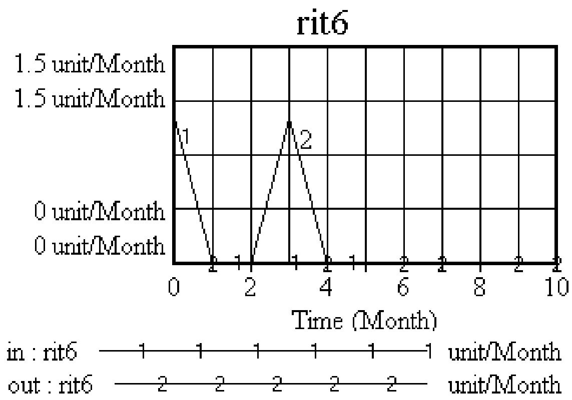

Figure 59: *Andamento dei flussi per il modello Vensim di figura 58 (dettagli nel testo)*

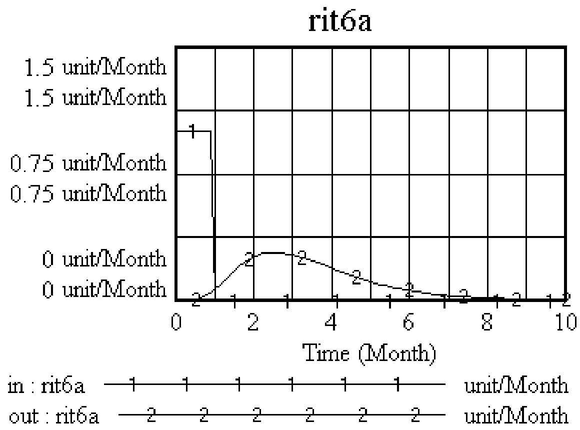

Figure 60: *Altri andamento dei flussi per il modello Vensim di figura 58 (dettagli nel testo)*

2.5 Un modello più complesso

Nella presente sezione si illustra un modello più complesso la cui struttura è riportata in figura 61 cui seguono le equazioni caratteristiche del modello.

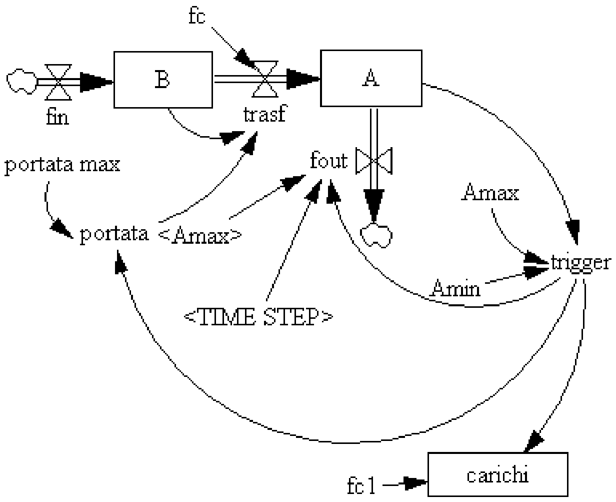

Figure 61: *Livelli connessi con svuotamento periodico (dettagli nel testo)*

```
(01)    A= INTEG (trasf-fout,0)
        Units: unit
(02)    Amax=250
        Units: unit
(03)    Amin=0
        Units: unit
(04)    B= INTEG (fin-trasf,0)
        Units: unit
(05)    carichi= INTEG (trigger*fc1,0)
        Units: Dmnl
(06)    fc=1
        Units: 1/Month
(07)    fc1=1
        Units: 1/Month
(08)    fin=1000
        Units: unit/Month [0,6000,10]
(09)    FINAL TIME  = 100
        Units: Month
        The final time for the simulation
(10)    fout=IF THEN ELSE(trigger=1, 0 ,Amax/TIME STEP)
        Units: unit/Month
```

```
(11)    INITIAL TIME  = 0
        Units: Month
        The initial time for the simulation.
(12)    portata=portata max *trigger
        Units: unit/Month
(13)    portata max=30
        Units: unit/Month [0,200,5]
(14)    SAVEPER=TIME STEP
        Units: Month [0,?]
        The frequency with which output is stored.
(15)    TIME STEP  = 0.0078125
        Units: Month [0,?]
        The time step for the simulation.
(16)    trasf=IF THEN ELSE( B*fc<portata , B*fc , portata )
        Units: unit/Month
(17)    trigger=IF THEN ELSE(A>=Amin :AND:A<=Amax,1 , 0 )
        Units: Dmnl
```

Lo scopo del modello è quello di descrivere un sistema caratterizzato da un serbatoio B in cui si accumulano liquidi ad un ritmo costante fin (equazione (08)). Tale serbatoio, di capacità teorica infinita, è collegato ad un altro serbatoio A di capacità massima A_{max} (equazione (02)) mediante un flusso $trasf$ (equazione (16)). Quando il secondo serbatoio è pieno è necessario svuotarlo per cui si deve bloccare il flusso in ingresso ad esso fino a svuotamento avvenuto.
Nel modello si suppone lo svuotamento istantaneo rispetto ai ritmi di riempimento (la scala temporale si misura in mesi). Lo svuotamento è pilotato da una variabile *trigger* (equazione (17)): se tale variabile vale 1 il flusso in uscita da A è nullo e, parallelamente si ha un trasferimento da B ad A, se tale variabile vale 0 si ha lo svuotamento di A e un blocco del trasferimento da B ad A. La variabile, di tipo livello, *carichi* tiene una contabilità del numero di volte in cui il serbatoio A viene svuotato. Gli andamenti di alcune delle variabili del modello sono dati in figura 62. Dalla figura si vede come:

1. il valore del livello A cresca da 0 a circa il valore massimo A_{max};

2. a quel punto *trigger* scatta a 0 e si ha un impulso sul flusso $fout$.

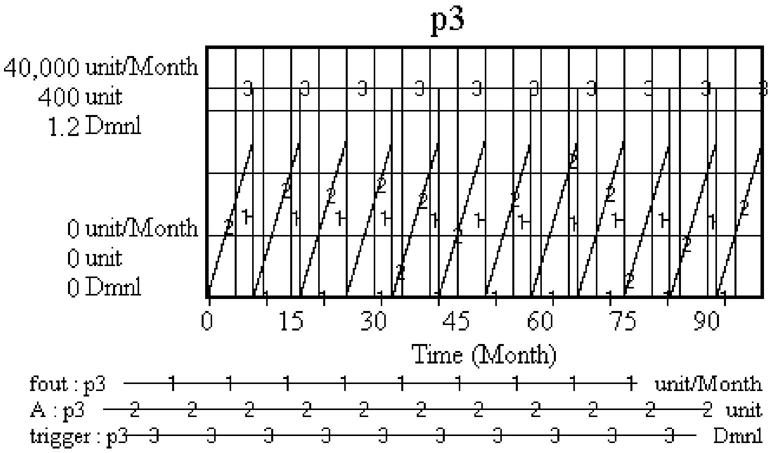

Figure 62: *Andamento di alcune grandezze del modello di figura 62 (dettagli nel testo)*

3 I modelli e la simulazione

3.1 Introduzione

Nella presente sezione[22] si inizia riprendendo alcuni dei concetti di base introdotti nelle precedenti sezioni per poi passare ad esaminare in modo più approfondito le problematiche relative alla esecuzione di simulazioni sui modelli dei sistemi dinamici. L'enfasi maggiore sarà posta sui DF ma, ogni volta sia necessario per dare maggiore chiarezza all'esposizione, faremo ricorso anche ai DCL. La parte iniziale della sezione deriva in gran parte da [Kir98].
Il metodo, che abbiamo delineato nelle sezioni precedenti, di un **approccio sistemico** (o del pensare in termini di sistemi, **sysetm thinking**) richiede che non vengano più considerati **eventi** isolati con le relative **cause** ma che si passi a considerare un insieme di elementi interagenti a formare un tutto che è stato chiamato **sistema**. Connessi al concetto di **sistema** ci sono i concetti di:

1. **interdipendenza** fra gli elementi del sistema;
2. **interazione** fra il sistema e gli elementi ad esso esterni che rappresentano il **mondo esterno**;
3. eventi generati dal mondo esterno che causano variazioni nel **comportamento** del sistema.

In base a questo approccio si parte dalla struttura interna di un sistema e si passa poi ad analizzare come questa sia influenzata da eventi nel mondo esterno. Il primo passo per arrivare a determinare la **struttura interna** di un sistema consiste nell'individuare i cosiddetti **schemi di comportamento** ovvero le modalità con cui le variabili caratteristiche del sistema cambiano nel tempo in modo da mostrare quelli che sono stati detti gli **andamenti di riferimento** o **caratteristici**. L'approccio sistemico si presenta come uno strumento per la risoluzione dei problemi nella misura in cui le strutture soggiacenti che determinano certi andamenti di riferimento sono note e le si può pensare presenti in tutti i casi in cui un certo andamento di riferimento si presenta. La soluzione consiste nell'agire sulla struttura in modo da eliminare un comportamento indesiderato e rafforzare uno desiderato a qualche scopo. Come abbiamo già avuto modo di vedere, gli andamenti di riferimento "tipici" cadono in una delle seguenti categorie, ad ognuna delle quali corrisponde una struttura che la genera:

1. una **crescita esponenziale** se una grandezza cresce ad un tasso di crescita crescente;
2. una **decrescita esponenziale** se una grandezza decresce ad un tasso di crescita crescente;

[22] Come è ovvio moltissimi saranno i richiami al materiale trattato nelle sezioni precedenti, richiami che saranno lasciati sottintesi tranne che nei casi in cui sia indispensabile la loro esplicitazione.

3. un andamento **asintotico** che tende ad un livello se una grandezza tende ad un livello costante partendo da un valore iniziale maggiore (decrescita) o minore (crescita);

4. un **un andamento a S** se una grandezza ha un andamento che può essere descritto da una crescita esponenziale seguita da un andamento asintotico in modo che la variabile si schiacci su un valore costante nel tempo;

5. un andamento di tipo **oscillatorio** se una grandezza assume valori attorno ad un valore costante con un andamento periodico.

Alcune combinazioni possibili sono le seguenti:

1. crescita esponenziale con oscillazioni sovrapposte;

2. decrescita esponenziale con oscillazioni sovrapposte;

3. andamento asintotico con sovrapposte oscillazioni;

4. andamento a S con sovrapposte oscillazioni.

3.2 Feedback e DCL: una rivisitazione

Per rappresentare la struttura dei sistemi si sono introdotte delle notazioni grafiche che permettono di tracciare sia i DCL sia i DF. I DCL permettono di rappresentare le relazioni esistenti fra gli elementi di un sistema mediante archi:

1. orientati,

2. con segno.

L'orientamento esprime una relazione di causa÷effetto fra due elementi e il segno descrive il tipo di relazione. Se si hanno due elementi collegati da un arco orientato nella forma:

$$A \longrightarrow_{+} B \tag{54}$$

si vuole significare che:

1. A è causa di B;

2. fra le due variabili esiste una relazione di proporzionalità diretta (inversa se il segno è $-$).

Il legame rappresentato dalla freccia si dice **legame causale**. Nel caso in cui, tramite una catena di legami causali, un elemento sia causa di sè stesso si dice che siamo in presenza di un **anello di feedback** o **anello causale**. In modo formale si ha che un **anello di feedback** è una catena chiusa di cause ed effetti che forma un cammino chiuso. Una catena di cause ed effetti che non forma un cammino chiuso si dice **anello aperto**. Ogni volta che si ragiona in termini di anelli aperti siamo al di fuori dell'approccio sistemico.
Dai segni sugli archi discende una caratterizzazione del segno o **tipo** degli anelli di feedback. Si hanno due casi:

1. il numero degli archi con segno $-$ è nullo o pari;
2. il numero degli archi con segno $-$ è dispari.

Nel primo caso l'anello si dice di tipo **positivo**, nel secondo si dice di tipo **negativo**.
A questo punto si stabilisce un legame fra la struttura interna di un sistema e gli andamenti di riferimento che abbiamo esaminato in precedenza. nella situazione più semplice si hanno i casi seguenti:

1. **anello di feedback positivo**;
2. **anello di feedback negativo**;
3. combinazione di un **anello di feedback positivo** e di un **anello di feedback negativo**;
4. **anello di feedback positivo** con un **ritardo**.

Nel primo caso si hanno grandezze in relazione fra di loro che mostrano andamenti crescenti nel tempo, andamenti di **crescita esponenziale**. In tal caso si ha una velocità di crescita crescente. Un anello di anello di feedback positivo è detto costituire un ciclo vizioso o virtuoso in funzione del tipo della crescita: nel primo caso la crescita è vista come un fenomeno negativo, nel secondo come un fenomeno positivo.
Nel secondo caso si hanno una o più grandezze che tendono ad un valore costante ovvero mostrano un **comportamento asintotico**. Ciò può accadere in due modi:

1. partendo da un valore inferiore ovvero con un andamento crescente;
2. partendo da un valore superiore ovvero con un andamento decrescente.

Gli anelli di feedback negativi danno stabilità ai sistemi (in senso positivo) oppure possono rappresentare una inerzia che si oppone alle variazioni (nei casi negativi).
Nel terzo caso si ha una combinazione di un **anello di feedback positivo** e di un **anello di feedback negativo**. In questo caso si ha un anello di feedback positivo, inizialmente dominante che causa una crescita esponenziale, cui subentra, dopo un certo tempo, un anello di feedback negativo che domina il comportamento del sistema e fa si che questo abbia un andamento di tipo asintotico.
Nel terzo caso si ha che un **anello di feedback positivo** con un **ritardo** può dare luogo ad oscillazioni dei seguenti tipi:

1. ad **ampiezza costante**;
2. ad **ampiezza decrescente** dette **oscillazioni smorzate**;
3. ad **ampiezza crescente**.

La presenza di oscillazioni nell'andamento di una o più variabili può essere associata alla presenza di ritardi lungo gli archi di un anello. Un ritardo fa si che una decisione venga presa sulla base di un valore non aggiornato di una variabile in modo che può risultare sia eccessiva sia insufficiente e ciò si traduce in oscillazioni attorno ad un valore costante.
Si fa notare che i processi di crescita esponenziale nei sistemi fisici reali prima o poi innescano fenomeni di "resistenza" che limitano la crescita e, al limite, la arrestano dal momento che causano l'insorgere, per cause varie, di anelli di feedback negativi.
A questo punto si danno alcune indicazioni di carattere generale su come si traccia un DCL. Il passo iniziale è quello di decidere quali sono gli **eventi** significativi per capire la **struttura di un sistema**. Dagli eventi si passa, anche solo in modo qualitativo, alla descrizione degli andamenti di riferimento delle grandezze significative. Fatto questo si usano le strutture viste degli anelli di reazione per costruire la **struttura iniziale** di un DCL che permetta di **spiegare** un dato andamento di riferimento.
Per procedere in modo più spedito si possono usare le seguenti regole pratiche:

1. gli elementi dei DCL sono le **variabili** di cui si vogliono determinare gli andamenti;

2. per ogni variabile non è necessario avere chiare sin da subito nè l'unità di misura nè la scala (valore minimo e valore massimo, passo di incremento) anche se una determinazione di tali fattori è, spesso, necessaria ai fini della simulazione;

3. i nomi delle variabili dovrebbero essere indicativi del ruolo che giocano nel sistema in relazione ai legami causali in cui sono coinvolte;

4. al momento di tracciare i legami causali è necessario valutare la presenza di effetti collaterali inattesi e decidere se ad essi debbano essere associati legami causali (magari introducendo altre variabili) o no;

5. nel caso di anelli causali di tipo negativo che prevedono la presenza di un livello di regime è bene che questo sia evidenziato mediante una variabile che segnala la distanza fra un valore corrente e tale valore finale;

6. se un legame causale fra due variabili risulta poco chiaro è possibile sostituirvi una catena di legami contenente variabili aggiuntive che svolgono un ruolo esplicativo;

7. è bene che un DCL sia tracciato in modo da essere il più semplice possibile compatibilmente con la necessità di rappresentare gli aspetti salienti del sistema di cui si stà costruendo il modello;

8. è sempre bene partire da DCL semplici e complicarli man mano in modo da ottenere comportamenti sempre più aderenti agli andamenti di riferimento.

3.3 DCL e DF: una rivisitazione

La figura 63 illustra un semplicissimo DCL. Si hanno tre variabili due delle quali sono collegate in un anello di feedback negativo. I nomi delle variabili sono stati scritti in modo che:

1. i nomi con iniziale maiuscola individuino livelli;
2. gli altri individuino flussi (o variabili ausiliarie).

Tale convenzione aiuta ad individuare il tipo di una variabile semplicemente considerandone il nome ma non è uno standard dei DCL. La figura 64, tenendo

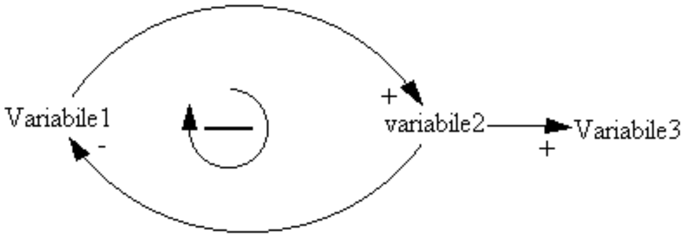

Figure 63: *Esempio di DCL (dettagli nel testo)*

conto delle convenzioni sui nomi delle variabili, rappresenta il DF corrispondente (a meno di una piccola modifica rappresentata dalla variabile esogena *tasso* che funge anche da fattore di conversione, vedi oltre). In tale figura è immediato individuare:

1. il livelli;
2. il flusso;
3. il flusso di materia;
4. il flusso di informazioni.

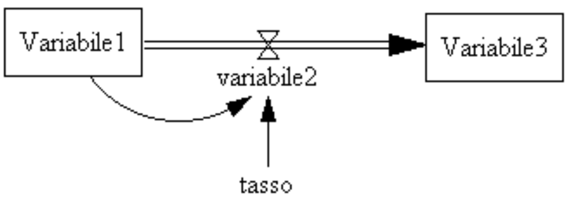

Figure 64: *DF corrispondente al DCL di figura 63 (dettagli nel testo)*

Tali elementi sono sufficienti per sviluppare un modello quantitativo che può essere usato per simulare il comportamento del modello. Si ricorda che:

1. un **livello** rappresenta l'accumulazione di qualcosa;
2. un **flusso** rappresenta una variazione in aumento (flusso entrante) o in diminuzione (flusso uscente) di un livello.

Ciò che interessa, nei modelli, sono le variazioni delle variabili nel tempo perchè in assenza di variazioni non è possibile parlare nè di **dinamica** nè di **controllo**. Oltre al flusso di materia in figura 64 e presente un flusso di informazione. I flussi di informazione rappresentano modalità di controllo non legate a trasferimenti di materia (che sono soggetti a vincoli di conservazione) in modo che si possono originare in ogni punto del modello e mettere in relazione fra di loro coppie di variabili qualunque.

3.4 La simulazione dei modelli

Perchè un DF ci permetta di capire come evolvono le diverse variabili nel tempo è necessario passare dalla rappresentazione grafica alle relazioni quantitative fra le variabili. Proseguendo l'esempio di figura 64, gli elementi quantitativi includono:

1. i valori iniziali del livelli;
2. le leggi che regolano gli andamenti dei flussi.

Nel momento in cui si decide di descrivere da un punto di vista quantitativo un modello si devono considerare vari fattori. Due fattori chiave sono:

1. il livello di dettaglio;
2. la gestione delle incertezze.

Di solito ci si pone ad un livello di dettaglio intermedio in relazione al problema e si escludono considerazioni di tipo probabilistico. Si deve tenere presente che di solito quello che si vuole è un modello relativamente semplice, di tipo quantitativo che dia risultati utili in tempi relativamente brevi se confrontati con l'urgenza del problema che si deve risolvere. A tale scopo si fanno le seguenti assunzioni semplificative ([Kir98]):

1. i flussi sono di tipo continuo ovvero la quantità che fluisce può essere divisa all'infinito rispetto sia al tempo sia alla quantità di materiale che fluisce;
2. i flussi non hanno componenti casuali in modo che un flusso è noto se lo sono le variabili che lo influenzano con archi di tipo informativo.

Una conseguenza di tali ipotesi è che si può usare una analogia di tipo idraulico per descrivere l'evoluzione dinamica di livelli e flussi interagenti in cui:

1. i livelli sono depositi di liquido,

2. i flussi sono valvole di immissione o scarico che controllano la velocità di scorrimento fra livelli.

In questo modo si ha che per specificare le equazioni descrittive di un modello servono:

1. i valori iniziali dei livelli;
2. le equazioni descrittive dei flussi.

Applicando quanto detto al caso di figura 64 si ottengono le equazioni di figura 65.

```
(1)   FINAL TIME  = 100
      Units: Month
      The final time for the simulation.
(2)   INITIAL TIME  = 0
      Units: Month
      The initial time for the simulation.
(3)   SAVEPER  =TIME STEP
      Units: Month [0,?]
      The frequency with which output is stored.
(4)   tasso=0.025
      Units: 1/Month
(5)   TIME STEP  = 1
      Units: Month [0,?]
      The time step for the simulation.
(6)   Variabile1= INTEG (-variabile2,100)
      Units: unita'
(7)   variabile2= tasso*max(0,Variabile1)
      Units: unita'/Month
(8)   Variabile3= INTEG (variabile2,0)
      Units: unita'
```

Figure 65: *Equazioni del DF di figura 64 (dettagli nel testo)*

Eseguendo la simulazione dall'istante iniziale all'istante finale (vedi le equazioni (1) e (2)) e risolvendo ogni $TIMESTEP$ le equazioni (6), (7) e (8) si ottengono gli andamenti riportati in figura 66 L'equazione (8) stabilisce che, istante per istante, $Variabile3$ contiene la "somma" totale dei contributi di $variabile2$ a partire da un valore iniziale 0 mentre per $Variabile1$ si ha una equazione che descrive uno "svuotamento" a partire da un valore iniziale pari a 100. In generale per un livello l'equazione descrittiva ha la forma seguente:

$$Livello(t) = Livello(0) + \int_0^t flusso\_in(\tau)\, d\tau - \int_0^t flusso\_out(\tau)\, d\tau \qquad (55)$$

ovvero il valore di *Livello* all'istante t coincide con il valore iniziale (in genere all'istante 0) cui si somma il contributo del flusso in ingresso ($\int_0^t flusso\_in(\tau) d\tau$) e si sottrae il contributo del flusso in uscita $\int_0^t flusso\_out(\tau)\, d\tau$. In genere i programmi per la simulazione di sistemi dinamici come Vensim sono in grado di determinare le equazioni descrittive dei livelli, a meno del valore iniziale e dell'unità di misura, mentre non sono in grado di stabilire la forma delle equazioni descrittive dei flussi. In figura 65 l'equazione (7) rappresenta un esempio di equazione

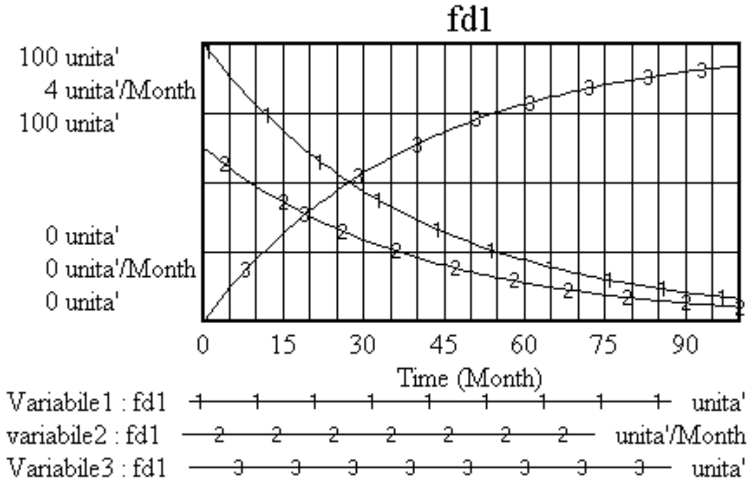

Figure 66: *Andamenti delle equazioni del DF di figura 65 (dettagli nel testo)*

che definisce l'andamento di un flusso nel tempo alla quale corrisponde il grafico etichettato come 2 in figura 66.
Risolvere le equazioni di figura 65 significa calcolare i valori delle variabili da $INITIALTIME$ a $FINALTIME$ ottenendo serie temporali in cui valori successivi sono calcolati a distanza di $TIME\,STEP$. Una volta ottenute le serie temporali le si può visualizzare sia variabile per variabile sia usando grafici contenenti più variabili e i relativi andamenti (vedi la figura 66). Il programma Vensim possiede molti comandi atti allo scopo. Per una loro trattazione si rimanda a [Ven05].
Una variante del modello di figura 64 è riportata in figura 67.

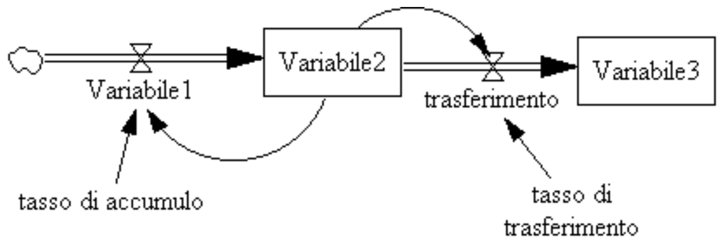

Figure 67: *Altro esempio di DF (dettagli nel testo)*

Gli andamenti delle variabili significative di tale modello sono riportati in figura 68 mentre in figura 69 sono riportate le equazioni relative. Le equazioni (4) e (5) definiscono le variabili $tasso\_di\_accumulo$ e $tasso\_di\_trasferimento$ come

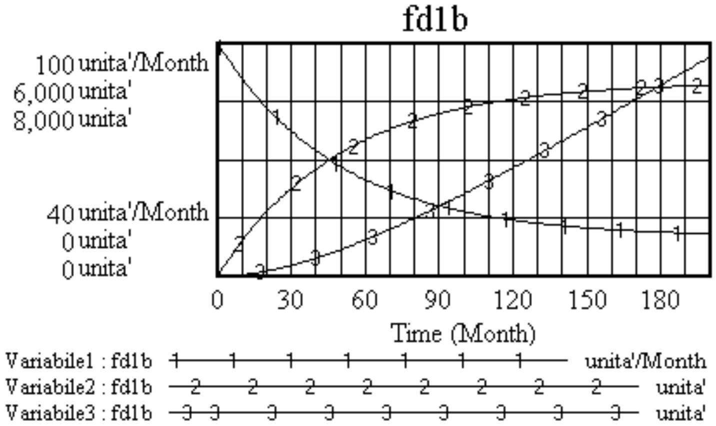

Figure 68: *Andamenti delle variabili del DF di figura 67 (dettagli nel testo)*

assumere valori compresi nell'intervallo $[0, 1]$, come accade in genere per variabili di tale tipo.
Da un esame delle equazioni di figura 69 è facile capire che, a partire dai valori iniziali dei livelli all'istante $INITIAL TIME$, vengono calcolati i valori dei flussi $Variabile1$ e $trasferimento$ e, sulla base di questi, si calcolano i valori successivi dei livelli e così via fino a che, di $TIMESTEP$ in $TIMESTEP$, non si arriva a $FINAL TIME$ quando la simulazione, se tutto va bene, termina. Possibili problemi sono:

1. una variabile che compare a denominatore in una espressione assume, in un qualche passo della simulazione, valore nullo;

2. una variabile che compare a numeratore in una espressione assume valori superiori al massimo rappresentabile (si ha **overflow**).

Dato un modello sotto forma sia di DCL sia di DF ai cui elementi costitutivi sono state associate delle equazioni e delle unità di misura, il programma Vensim possiede, infatti, strumenti formali per verificare la correttezza formale sia delle equazioni sia delle unità di misura mentre non ne ha, come è intuitivo, per prevedere ed, eventualmente, impedire l'insorgere di tali condizioni di errore a tempo di esecuzione a meno che non si usino accorgimenti ad hoc (quali l'uso di funzioni di MIN o MAX). Per ulteriori dettagli si rimanda a [Ven05].
Nelle equazioni delle figure 69 e 65 le variabili esogene sono ipotizzate avere valori costanti nel tempo. Nel caso di figura 69 le variabili esogene $tasso\_di\_accumulo$ e $tasso\_di\_trasferimento$ hanno un valore costante pari a 0.01. Un modo per verificare la validità di un modello è quello di applicare perturbazioni alle variabili esogene e vedere come tali perturbazioni influenzano gli andamenti delle variabili endogene, in questo caso $Variabile1$, $Variabile2$ e $Variabile2$. In genere si può procedere in due modi:

```
(01)    FINAL TIME  = 200
        Units: Month
        The final time for the simulation.
(02)    INITIAL TIME  = 0
        Units: Month
        The initial time for the simulation.
(03)    SAVEPER  =TIME STEP
        Units: Month [0,?]
        The frequency with which output is stored.
(04)    tasso di accumulo=0.01
        Units: 1/Month [0,1]
(05)    tasso di trasferimento=0.01
        Units: 1/Month [0,1]
(06)    TIME STEP  = 1
        Units: Month [0,?]
        The time step for the simulation.
(07)    trasferimento= Variabile2*tasso di trasferimento
        Units: unita'/Month
(08)    Variabile1=100-Variabile2*tasso di accumulo
        Units: unita'/Month
(09)    Variabile2= INTEG (Variabile1-trasferimento,0)
        Units: unita'
(10)    Variabile3= INTEG (trasferimento,0)
        Units: unita'
```

Figure 69: *Equazioni del DF di figura 67 (dettagli nel testo)*

1. si usano funzioni dette di test per imporre ad una variabile esogena una variazione di un qualche tipo in un qualche istante di tempo;
2. si esegue la simulazione partendo dai valori costanti delle variabili esogene e li si modifica a tempo di esecuzione in modo da causare variazioni degli andamenti delle variabili esogene.

Come di consueto vedremo le due metodologie nel caso di Vensim rimandando a [Ven05] per i dettagli operativi.
Alcune delle funzioni di test disponibili in Vensim e applicabili alle variabili esogene sono le seguenti (vedi la figura 70):

1. funzione **impulso unitario** $PULSE$ di ampiezza pari a 1 di cui si devono specificare **istante di inizio** t_i e durata T (caso (a));
2. funzione **serie di impulsi unitari** $PULSE\,TRAIN$ di ampiezza pari a 1 di cui si devono specificare **istante di inizio** t_i, durata T, **periodo di ripetizione** T_p e istante di fine t_f (caso (b));
3. funzione **rampa** $RAMP$ di cui si devono specificare la **pendenza** e gli istanti di inizio t_i e di fine t_f (caso (c));
4. funzione **gradino** $STEP$ di cui si devono specificare l'**altezza** M e l'istante di inizio t_0 (caso (d)).

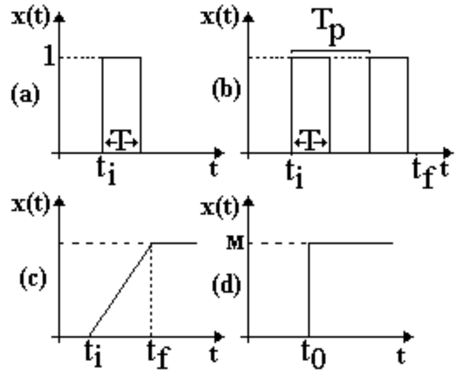

Figure 70: *Funzioni test (dettagli nel testo)*

Un esempio di applicazione di una funzione di test lo si ha se si applica un disturbo di tipo gradino sulla variabile *tasso_di_accumulo* in modo che l'espressione diventi:

$$tasso\_di\_accumulo = 0.01 + STEP(0.01, 30) \quad (56)$$

Eseguendo la simulazione si ottengono gli andamenti di figura 71. Un esem-

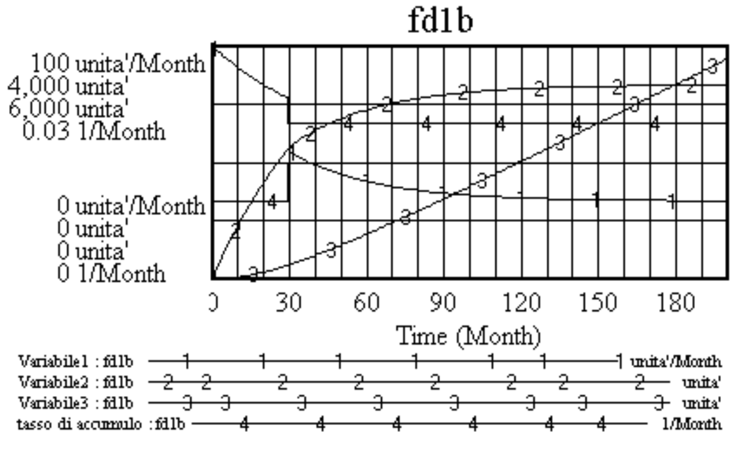

Figure 71: *Perturbazione di una variabile esogena (dettagli nel testo)*

pio di esecuzione di modifiche a tempo di esecuzione lo si ha in figura 72. In figura 72 sono rappresentati due cursori sotto le due variabili esogene. Agendo su tali cursori si modifica il valore corrente di ognuna delle variabili in modo da variare gli andamenti delle variabili endogene che sono mostrati in sovrimpressione alle variabili stesse. Nel caso presentato si agisce sulla variabile

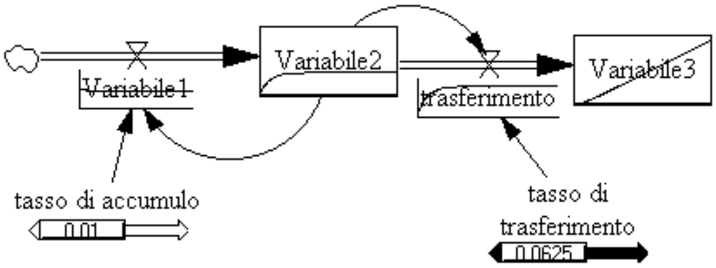

Figure 72: *Variazioni a tempo di esecuzione(dettagli nel testo)*

$tasso\_di\_trasferimento$. Si fa notare che i due approcci sono radicalmente diversi in quanto:

(1) il primo permette di imporre alle singole variabili esogene particolari andamenti nel tempo in grado di simulare variazioni dell'influenza del mondo esterno sul modello,

(2) il secondo consiste nel variare i valori delle variabili esogene, valori che rimangono costanti dal momento della variazione in poi.

Nel caso (1) si possono combinare fra loro le varie funzioni ad istanti diversi in modo da ottenere andamenti complessi.
Nel caso (2) si agisce come se si applicassero delle funzioni a gradino alle singole variabili esogene ma senza aver nessun controllo sugli istanti di applicazione. In questo modo non è possibile valutare l'effetto, ad esempio, di una variazione a gradino applicata ad una variabile esogena in T_0 con un'altra variazione a gradino applicata ad un'altra variabile esogena in $T_1 > T_0$. In più non si ha la possibilità di applicare funzioni di altro tipo nè combinazioni di funzioni.

4 Equazioni differenziali e relativi modelli

4.1 Introduzione

Nella presente sezione si introducono alcune equazioni differenziali tipiche del primo ordine e di ordine superiore, che diremo paradigmatiche, a ciascuna delle quali si associa il corrispondente modello ottenuto utilizzando il programma Vensim.
Vedremo inizialmente equazioni differenziali del primo ordine per poi passare a equazioni differenziali di ordine superiore ottenute mediante sistemi di due o più equazioni differenziali. Verranno, infine, presentati alcuni modelli nei quali la presenza di limitazioni di tipo fisico pone dei vincoli alla crescita di una o più variabili caratteristiche.
Di ogni modello verranno presentate simulazioni utilizzando conoscenze a questo punto acquisite.

4.2 Equazioni differenziali del primo ordine

Il punto di partenza è la seguente equazione differenziale lineare del primo ordine[23]:

$$\frac{dX}{dt} = aX \tag{57}$$

con valore iniziale X_0 (per $t = 0$) e $a \in \mathbb{R}$. La (57) è facilmente risolvibile (per i dettagli vedi l'Appendice A) ovvero se ne può dare la seguente soluzione in forma chiusa:

$$X(t) = X_0 e^{at} \tag{58}$$

Nella (57) e nella (58) il parametro a rappresenta l'inverso di una costante di tempo:

$$\tau = \frac{1}{a} \tag{59}$$

e il suo segno determina l'andamento crescente ($a > 0$) o decrescente ($a < 0$) della $X(t)$ a partire dal valore iniziale X_0. La (57) può essere facilmente tradotta in un diagramma Vensim (vedi la Figura 73). Il modello Vensim corrispondente alla (57) lo si ricava passando alla corrispondente equazione alle differenze finite ovvero scrivendo:

$$\frac{\Delta X}{\Delta t} = aX(t) \tag{60}$$

ovvero:

$$\frac{X(t + \Delta t) - X(t)}{\Delta t} = aX(t) \tag{61}$$

da cui:

$$X(t + \Delta t) = X(t) + aX(t)\Delta t \tag{62}$$

[23]Nei diagrammi Vensim useremo nomi di variabili come $dotX$ per indicare $\frac{dX}{dt}$ e $X_0 = X0$ per indicare il valore iniziale della variabile X. Saranno usate altre semplificazioni chiarite sul momento ma, di solito, facilmente comprensibili. Inoltre faremo uso della scrittura X' per indicare $\frac{dX}{dt}$, X'' per la derivata seconda, X''' per la derivata terza e così via.

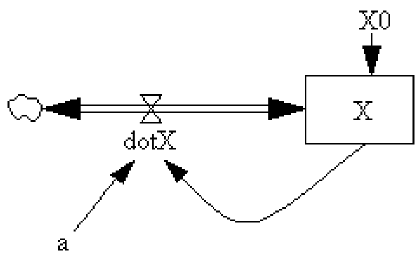

Figure 73: *Diagramma Vensim della (57) (dettagli nel testo)*

Definendo una variabile di tipo flusso[24] $dotX = Y(t) = aX(t)$ si ottiene il sistema di equazioni seguente:

$$\begin{cases} X(t+\Delta t) = X(t) + Y(t)\Delta t \\ Y(t) = aX(t) \end{cases} \tag{63}$$

cui corrisponde il modello Vensim di Figura 73.
Il procedimento visto ha validità generale e sarà applicato ripetutamente in queste note, sia pure non sempre con questo dettaglio.
Qui di seguito si riportano le equazioni associate a tale modello.

```
(1) a=0.1
    Units: 1/Month [-1,1,0.01]
(2) dotX=a*X
    Units: unit/Month
(3) FINAL TIME  = 100
    Units: Month
    The final time for the simulation.
(4) INITIAL TIME  = 0
    Units: Month
    The initial time for the simulation.
(5) SAVEPER  = TIME STEP
    Units: Month [0,?]
    The frequency with which output is stored.
(6) TIME STEP  = 1
    Units: Month [0,?]
    The time step for the simulation.
(7) X= INTEG (dotX,X0)
    Units: unit
(8) X0=1
    Units: unit
```

[24]La $Y(t)$ è una variabile di tipo flusso dato che dimensionalmente è data dal rapporto fra $[X]$ e $[\Delta t]$ dove con $[a]$ si indica l'unità di misura della grandezza a.

Equazioni Vensim della 73 (dettagli nel testo)

Delle equazioni su elencate si fanno notare:

1. la (2) che traduce la (57)
2. la (7) che, in un certo senso, corrisponde alla (58).

Le altre equazioni definiscono:

1. le grandezze indispensabili per la simulazione (quali $FINAL\,TIME$, $INITIAL\,TIME$, $SAVEPER$ e $TIME\,STEP$),
2. il parametro (a) e il valore iniziale del livello, ovvero $X0$.

La prima modifica che si può apportare alla (57) è quella di aggiungere un termine costante $b \in \mathbb{R}$:

$$\frac{dX}{dt} = aX + b \tag{64}$$

La (64) ha una soluzione in forma chiusa (per i dettagli vedi l'Appendice A) ottenuta come somma della soluzione dell'equazione omogenea associata e di una soluzione particolare $X(t) = cost$ della (64). Tale soluzione ha la forma seguente:

$$X(t) = (X_0 + \frac{b}{a})e^{at} - \frac{b}{a} \tag{65}$$

Se $b = 0$ la (65) si riduce alla (58). Per $t = 0$ la (65) dà l'identità $X(0) = X_0$. Come è ovvio, a noi non interessano tanto i procedimenti analitici di soluzione (che danno le soluzioni in forma chiusa e che non sono disponibili per ogni tipo di equazione differenziale) piuttosto interessa stabilire i modelli corrispondenti alle singole equazioni differenziali su cui eseguire le simulazioni che ci forniscono per via numerica le soluzioni cercate. Nel seguito, pertanto, ci limiteremo a:

1. presentare il problema differenziale,
2. presentare il corrispondente modello,
3. illustrarne le equazioni chiave,
4. presentare alcune simulazioni.

In tale ottica le figure che seguono contengono, nell'ordine:

1. il modello Vensim della (64);
2. le equazioni descrittive del modello;
3. una simulazione del modello per $a = 0$ e $b > 0$;
4. una simulazione del modello per $a < 0$ e $b = 0$;
5. una simulazione del modello per $a < 0$ e $b > 0$.

In merito alla (64) si fa notare come si possa ricavare il modello Vensim corrispondente di Figura 74 mediante le seguenti manipolazioni.
Per prima cosa si scrive la corrispondente equazione alle differenze finite:

$$\frac{\Delta X}{\Delta t} = aX(t) + b \tag{66}$$

da cui si ha:

$$\frac{X(t+\Delta t) - X(t)}{\Delta t} = aX(t) + b \tag{67}$$

ovvero:

$$X(t+\Delta t) = X(t) + (aX(t) + b)\Delta t \tag{68}$$

Definendo una variabile di tipo flusso $Y(t) = aX(t) + b$ si ottiene il sistema di equazioni seguente:

$$\begin{cases} X(t+\Delta t) = X(t) + Y(t)\Delta t \\ Y(t) = aX(t) + b \end{cases} \tag{69}$$

cui corrisponde il modello Vensim di Figura 74.

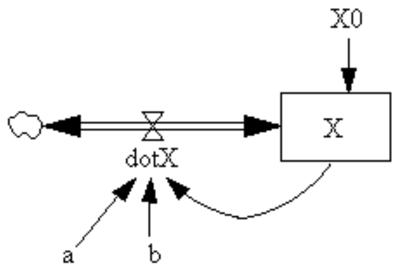

Figure 74: *Diagramma Vensim della (64) (dettagli nel testo)*

Qui di seguito sono riportate le equazioni associate al modello di Figura 74.

```
(01)    a=-1
        Units: 1/Month [-1,1,0.1]
(02)    b=2
        Units: unit/Month [-10,10]
(03)    dotX=a*X+b
        Units: unit/Month
(04)    FINAL TIME  = 10
        Units: Month
        The final time for the simulation.
(05)    INITIAL TIME  = 0
        Units: Month
        The initial time for the simulation.
(06)    SAVEPER  = TIME STEP
```

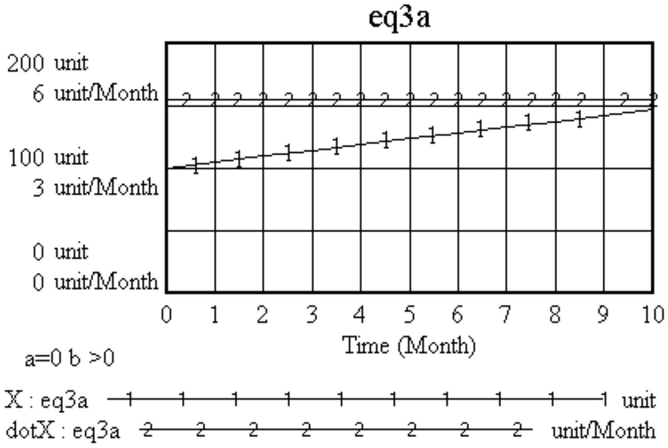

Figure 75: *Prima simulazione del modello di Figura 74 (dettagli nel testo)*

```
        Units: Month [0,?]
        The frequency with which output is stored.
(07)    TIME STEP  = 0.125
        Units: Month [0,?]
        The time step for the simulation.
(08)    X= INTEG (dotX,X0)
        Units: unit
(09)    X0=100
        Units: unit
```

In relazione alle equazioni descrittive del modello ci preme di segnalare:

1. la (02) che definisce b;
2. la (03) che definisce la (64).

Nell'eseguire la terza simulazione si è modificato il valore di $TIME\_STEP$ impostandolo pari a 0.125.
A questo punto si può ulteriormente complicare la (64) definendo b come variabile nel tempo con una legge esponenziale. In tal modo si ha:

$$\frac{dX}{dt} = aX(t) + b(t) \tag{70}$$

con valore iniziale X_0 e:

$$b(t) = b_0 e^{\frac{t}{tc}} \tag{71}$$

in cui $tc = \tau$ è la costante di tempo e b_0 è un valore iniziale.
La Figura 78 contiene il modello Vensim della (70)[25] le cui equazioni descrittive sono riportate qui di seguito:

[25] Il modello lo si ricava trasformando l'equazione differenziale data nella corrispondente equazione alle differenze ed eseguendo le trasformazioni usuali.

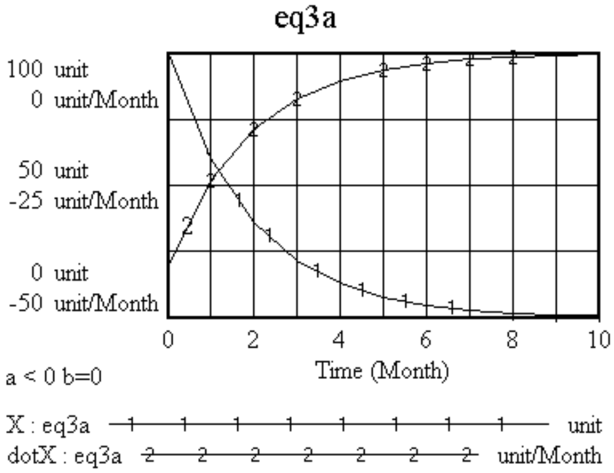

Figure 76: *Seconda simulazione del modello di Figura 74 (dettagli nel testo)*

```
(01)    a=0
        Units: 1/Month [-1,1,0.1]
(02)    b=b0*EXP( Time/tc )
        Units: unit/Month
(03)    b0=2
        Units: unit/Month
(04)    dotX=a*X+b
        Units: unit/Month
(05)    FINAL TIME  = 10
        Units: Month
        The final time for the simulation.
(06)    INITIAL TIME  = 0
        Units: Month
        The initial time for the simulation.
(07)    SAVEPER=TIME STEP
        Units: Month [0,?]
        The frequency with which output is stored.
(08)    tc=10
        Units: Month
(09)    TIME STEP  = 1
        Units: Month [0,?]
        The time step for the simulation.
(10)    X= INTEG (dotX,X0)
        Units: unit
(11)    X0=100
        Units: unit
```

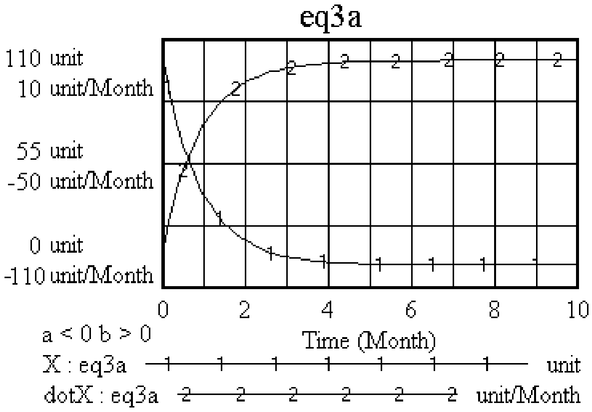

Figure 77: *Terza simulazione del modello di Figura 74 (dettagli nel testo)*

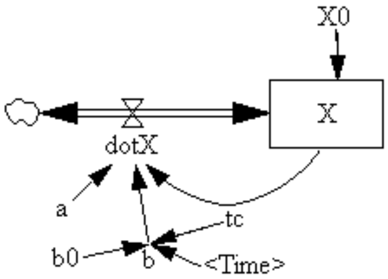

Figure 78: *Diagramma Vensim della (70) (dettagli nel testo)*

Il parametro tempo t che compare nell'esponenziale è ottenuto utilizzando la variabile $< Time >$ che contiene il conteggio dei clock usati durante le simulazioni: tale variabile rappresenta l'avanzare del tempo interno a Vensim.
L'equazione significativa è la (02) (e le equazioni "collegate" ovvero la (03) e la (08)): tale equazione definisce la funzione $b(t)$ mentre le altre sono identiche a quelle del modello di Figura 74.
La Figura 79 presenta gli andamenti di alcune variabili del modello Vensim di Figura 78 nel caso in cui:

1. $a = 0$,
2. $b_0 = 2$,
3. $X_0 = 100$.

La Figura 80 presenta gli andamenti di alcune variabili del modello Vensim di Figura 78 nel caso in cui:

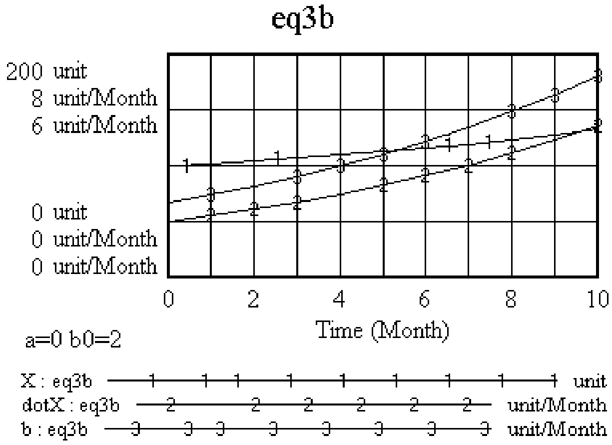

Figure 79: *Prima simulazione del modello Vensim di Figura 78 (dettagli nel testo)*

1. $a = -0.3$,
2. $b_0 = -3$,
3. $X_0 = 100$.

4.3 Sistemi di equazioni differenziali ed equazioni differenziali di ordine superiore al primo

In molti casi sorge la necessità di risolvere equazioni differenziali di ordine superiore al primo in una variabile, complete o meno. Se si vuole utilizzare Vensim, dal momento che il programma, per sua natura, consente di impostare solo equazioni differenziali del primo ordine, l'unica soluzione è quella di trasformare una equazione differenziale di ordine superiore al primo in una variabile in un sistema di più equazioni differenziali del primo ordine, facendo uso di variabili ausiliarie e di tecniche di trasformazione che facciano uso, qualora necessario, di operazioni di derivazione.
In quanto segue, data una equazione differenziale di un certo tipo di ordine superiore al primo, ci limiteremo, pertanto a:

1. trasformarla in un sistema equivalente di equazioni differenziali del primo ordine;
2. darne il modello Vensim corrispondente;
3. eseguire sul modello alcune simulazioni.

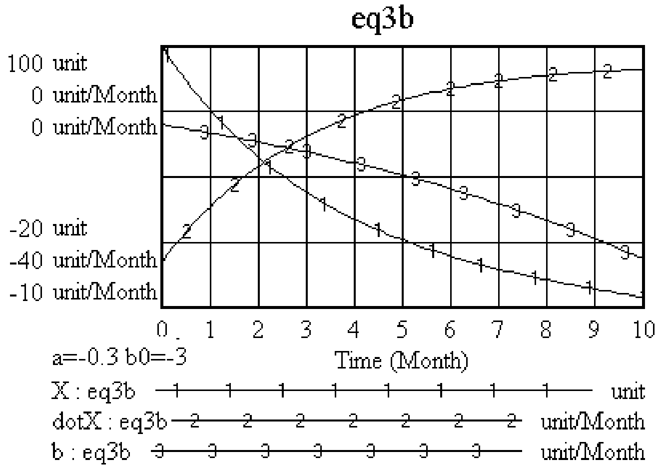

Figure 80: *Seconda simulazione del modello Vensim di Figura 78 (dettagli nel testo)*

Lo schema che seguiremo sarà sostanzialmente il seguente. Si imposta una equazione differenziale completa di un ordine superiore al primo, ovvero di ordine $n > 1$, la si trasforma in un sistema di n equazioni differenziali del primo ordine, si traccia il modello Vensim corrispondente a tale sistema e su di esso si eseguono alcune simulazioni.

4.3.1 Equazioni differenziali del secondo ordine

Si suppone di avere:

$$Y'' = aY' + bY + c \tag{72}$$

La (72) permette di ottenere tutte le varianti possibili agendo sui valori dei parametri a, b, c. In quanto segue supporremo che tali parametri assumano valori in $\mathbb{R} \setminus \{0\}$.
Se si vuole usare Vensim per risolvere una equazione differenziale di tale tipo è necessario trasformarla in un insieme di equazioni differenziali del primo ordine per poi passare, per ciascuna di esse, alla corrispondente equazione alle differenze finite e poi arrivare alle equazioni in termini di livelli e flussi.
In relazione alla (72) si procede definendo una variabile ausiliaria:

$$Z = Y' \tag{73}$$

in modo da scrivere la (72) sotto forma di un sistema di equazioni differenziali del primo ordine:

$$\begin{cases} Z'(t) = aZ(t) + bY(t) + c \\ Y'(t) = Z(t) \end{cases} \tag{74}$$

Passando dalle (74) alle corrispondenti equazioni alle differenze finite ed eseguendo alcuni semplici calcoli si arriva alle seguenti relazioni:

$$\begin{cases} Z(t+\Delta\, t) = Z(t) + (aZ(t) + bY(t) + c)\Delta\, t \\ Y(t+\Delta\, t) = Y(t) + Z(t)\Delta\, t \end{cases} \tag{75}$$

A questo punto, per semplicità, si può introdurre una variabile ausiliaria:

$$W(t) = aZ(t) + bY(t) + c \tag{76}$$

in modo che le (75) diventino:

$$\begin{cases} Z(t+\Delta\, t) = Z(t) + W(t)\Delta\, t \\ Y(t+\Delta\, t) = Y(t) + Z(t)\Delta\, t \\ W(t) = aZ(t) + bY(t) + c \end{cases} \tag{77}$$

Nelle (77) il $\Delta\, t$ rappresenta l'incremento finito di tempo fra due valutazioni successive delle grandezze che, negli esempi visti finora, abbiamo denotato con $TIME\_STEP$.
A questo punto è buona norma eseguire una analisi dimensionale delle (77). Da tale analisi si vede come, volendo che Y e Z siano viste come livelli con unità di misura generica *unit* mentre W sia dimensionalmente un flusso, si deve avere:

$$\begin{cases} [\mathrm{a}] = \frac{1}{t} \\ \\ [\mathrm{b}] = \frac{1}{t} \\ \\ [\mathrm{c}] = \frac{unit}{t} \end{cases} \tag{78}$$

Date le (78), la prima delle (77) risulta dimensionalmente a posto mentre nella seconda serve un fattore di conversione d di valore unitario e avente la dimensione di una frequenza (ovvero l'inverso di un tempo). Le relazioni da cui ricavare il modello Vensim sono, pertanto, le seguenti:

$$\begin{cases} Z(t+\Delta\, t) = Z(t) + W(t)\Delta\, t \\ Y(t+\Delta\, t) = Y(t) + d * Z(t)\Delta\, t \\ W(t) = aZ(t) + bY(t) + c \end{cases} \tag{79}$$

A questo punto non resta che:

1. definire il modello Vensim che implementa le (79);

2. eseguire alcune simulazioni.

Si fa presente come, rilassando il vincolo $a, b, c \in \mathbb{R} \setminus \{0\}$ si possano ottenere tutte le varianti possibili della (72). L'analisi dettagliata di tali varianti esula dallo scopo delle presenti note. Una analisi quantitative può essere effettuata facilmente agendo sulle variabili del modello Vensim e ponendo, di volta in volta, a 0 uno o più di tali parametri.

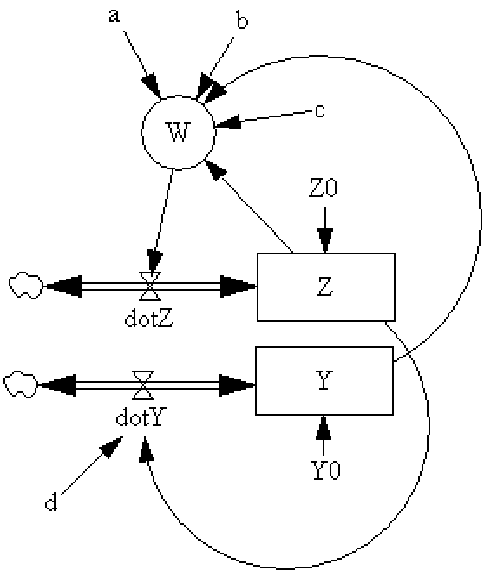

Figure 81: *Modello Vensim delle (79) (dettagli nel testo)*

In Figura 81 si riporta il modello Vensim descritto dalle (79).
Per semplicità si sono introdotte due variabili ausiliarie *dotY* e *dotZ* il cui ruolo risulta chiaro dalla analisi delle equazioni descrittive del modello riportate qui di seguito:

```
(01)    a=0.2
        Units: 1/Month
(02)    b=-0.3
        Units: 1/Month
(03)    c=-2
        Units: unit/Month
(04)    d=1
        Units: 1/Month [1,1,0.1]
(05)    dotY=d*Z
        Units: unit/Month
(06)    dotZ=W
        Units: unit/Month
(07)    FINAL TIME  = 12
        Units: Month
        The final time for the simulation.
(08)    INITIAL TIME  = 0
        Units: Month
        The initial time for the simulation.
(09)    SAVEPER  = TIME STEP
```

```
        Units: Month [0,?]
        The frequency with which output is stored.
(10)    TIME STEP  = 1
        Units: Month [0,?]
        The time step for the simulation.
(11)    W=a*Z+b*Y+c
        Units: unit/Month
(12)    Y= INTEG (dotY,Y0)
        Units: unit
(13)    Y0=2
        Units: unit
(14)    Z= INTEG (dotZ,Z0)
        Units: unit
(15)    Z0=1
        Units: unit
```

Con i valori dei parametri a, b, c dati dalle equazioni $(01), (02), (03)$, eseguendo una simulazione, si ottengono gli andamenti di Figura 82.

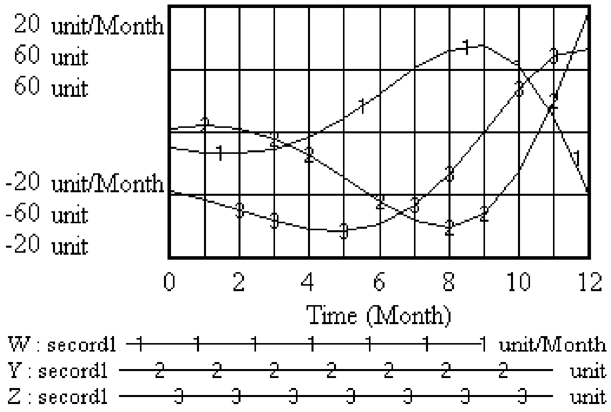

Figure 82: *Andamenti delle variabili del modello di Figura 81 (dettagli nel testo)*

Per avere una idea più chiara degli andamenti delle variabili monitorate è necessario eseguire la simulazione su un tempo più lungo. Eseguendo la simulazione su un tempo doppio del precedente si ottengono i risultati di Figura 83.
La Figura 84 riporta il risultato di una simulazione nel caso in cui sia:

1. $a = -0.35$
2. $b = -0.3$
3. $c = 0$

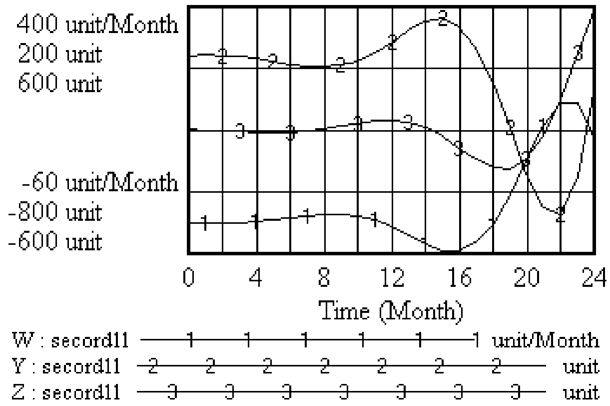

Figure 83: *Altri andamenti delle variabili del modello di Figura 81 (dettagli nel testo)*

in modo che la (72) ammetta solo la soluzione dell'equazione omogenea associata che risulta una oscillazione smorzata, essendo i coefficienti a e b di segno negativo.

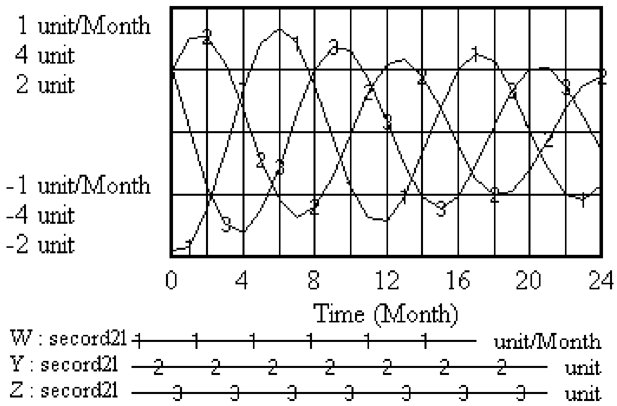

Figure 84: *Altri andamenti delle variabili del modello di Figura 81 (dettagli nel testo)*

4.3.2 Equazioni differenziali del terzo ordine e di ordine superiore

In questa sezione si esamina il caso della seguente equazione differenziale completa del terzo ordine:

$$Y''' = aY'' + bY' + cY + d \tag{80}$$

È facile vedere come la (80) possa essere trasformata nel seguente sistema di tre equazioni differenziali del primo ordine:

$$\begin{cases} Y'(t) = X(t) \\ X'(t) = Z(t) \\ Z'(t) = aZ(t) + bX(t) + cY(t) + d \end{cases} \tag{81}$$

Passando alle corrispondenti equazioni alle differenze finite si ottiene il seguente sistema:

$$\begin{cases} Y(t + \Delta\, t) = Y(t) + X(t)\Delta\, t \\ X(t + \Delta\, t) = X(t) + Z(t)\Delta\, t \\ Z(t + \Delta\, t) = Z(t) + (aZ(t) + bX(t) + cY(t) + d)\Delta\, t \end{cases} \tag{82}$$

Analogamente a quanto fatto nel caso precedente della equazione del secondo ordine, si può introdurre una variabile ausiliaria ed ottenere il seguente sistema di relazioni cui si associa il modello Vensim su cui eseguiremo alcune simulazioni:

$$\begin{cases} Y(t + \Delta\, t) = Y(t) + X(t)\Delta\, t \\ X(t + \Delta\, t) = X(t) + Z(t)\Delta\, t \\ Z(t + \Delta\, t) = Z(t) + W(t)\Delta\, t \\ W(t) = aZ(t) + bX(t) + cY(t) + d \end{cases} \tag{83}$$

Le (83) vanno aggiustate dimensionalmente se si vuole che le X, Y e Z siano dei livelli e la W un flusso[26]. A tale scopo si devono introdurre dei coefficienti dimensionali a valore unitario per i quali si rimanda al modello Vensim e alle relative equazioni.
Il modello corrispondente alle relazioni (83) è riportato in Figura 85. Le equazioni associate al modello di Figura 85 sono le seguenti:

```
(01)    a=0.1
        Units: 1/Month
(02)    b=-0.2
        Units: 1/Month
(03)    c=0.3
        Units: 1/Month
(04)    d=2
        Units: unit/Month
(05)    d1=1
        Units: 1/Month [1,1,0.1]
```

[26]Si ricorda che se un livello L ha $[L] = unit$ un flusso ad esso associato F ha $[F] = unit/t$.

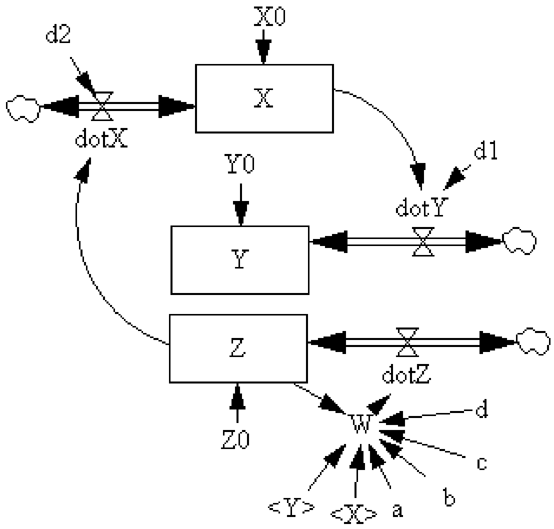

Figure 85: *Modello Vensim delle (83) (dettagli nel testo)*

```
(06)    d2=1
        Units: 1/Month [1,1,0.1]
(07)    dotX=d2*Z
        Units: unit/Month
(08)    dotY=d1*X
        Units: unit/Month
(09)    dotZ=a*Z+b*X+c*Y+d
        Units: unit/Month
(10)    FINAL TIME  = 100
        Units: Month
        The final time for the simulation.
(11)    INITIAL TIME  = 0
        Units: Month
        The initial time for the simulation.
(12)    SAVEPER  = TIME STEP
        Units: Month [0,?]
        The frequency with which output is stored.
(13)    TIME STEP  = 1
        Units: Month [0,?]
        The time step for the simulation.
(14)    X= INTEG (dotX,X0)
        Units: unit
(15)    X0=0.1
        Units: unit
```

```
(16)      Y= INTEG (dotY,Y0)
          Units: unit
(17)      Y0=1
          Units: unit
(18)      Z= INTEG (dotZ,Z0)
          Units: unit
(19)      Z0=1.2
          Units: unit
```

La figura 86 riporta il risultato di una simulazione del modello di Figura 85 per i seguenti valori dei parametri:

1. $a = -1$
2. $b = -0.3$
3. $c = -0.1$
4. $d = 2.25$

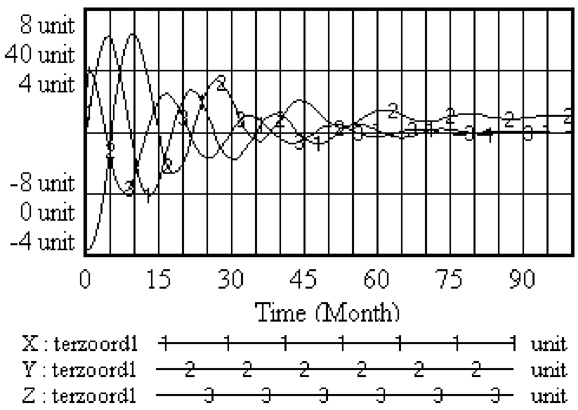

Figure 86: *Altri andamenti delle variabili del modello di Figura 85 (dettagli nel testo)*

Prima di accennare brevemente alle equazioni differenziali di ordine n si presenta una equazione differenziale di terzo ordine in forma generica:

$$Y''' = a_3 Y'' + a_2 Y' + a_1 Y + a_0 \tag{84}$$

Dalla (84) si può passare al seguente sistema di tre equazioni differenziali del primo ordine:

$$\begin{cases} Y'(t) = X_2(t) \\ X_2'(t) = x_1(t) \\ X_1'(t) = a_3 X_1 + a_2 X_2 + a_1 Y + a_0 \end{cases} \tag{85}$$

in cui i coefficienti a_i hanno le opportune dimensioni. Nel passaggio alle equazioni alle differenze finite, in questo caso come nel caso generico, sarà poi necessario introdurre coefficienti dimensionali a valore unitario per armonizzare le dimensioni di livelli e flussi che compaiono nelle varie equazioni, il tutto secondo procedimenti visti più volte e ormai noti.
Nel caso delle equazioni differenziali di ordine n nella forma seguente:

$$Y^n = a_n Y^{n-1} + a_{n-1} Y^{n-2} \dots a_2 Y' + a_1 Y + a_0 \tag{86}$$

Dalle (86) si può passare facilmente al seguente sistema di n equazioni differenziali del primo ordine, per il quale valgono, mutatis mutandis, considerazioni analoghe a quelle fatte per le (85):

$$\begin{cases} X_1'(t) = a_n X_1 + a_{n-1} X_2 + \dots + a_2 X_{n-1} + a_1 Y + a_0 \\ Y' = X_{n-1} \\ X_{n-1}' = X_{n-2} \\ \vdots \\ X_2' = X_1 \end{cases} \tag{87}$$

Nelle presenti note non esamineremo in ulteriore dettaglio equazioni differenziali di ordine superiore al terzo e i rispettivi modelli.

4.4 Altre equazioni differenziali

Nella presente sezione si esaminano brevemente i modelli Vensim di equazioni differenziali del tipo:

$$X'(t) = aX^2(t) \tag{88}$$

La (88) può essere riscritta come, passando alle differenze finite:

$$\frac{\Delta\, X(t)}{\Delta\, t} = aX^2(t) \tag{89}$$

ovvero come, considerando che $\Delta\, X(t) = X(t + \Delta\, t) - X(t)$:

$$X(t + \Delta\, t) = X(t) + aX^2(t)\Delta\, t \tag{90}$$

per arrivare al seguente sistema di equazioni:

$$\begin{cases} X(t + \Delta\, t) = X(t) + Y(t)\Delta\, t \\ Y(t) = aX^2(t) \end{cases} \tag{91}$$

in cui la X individua un livello e la Y un flusso.
Alle (91) corrisponde il modello Vensim di Figura 87 in cui la variabile $Y(t)$ è denotata (come in altri casi) come $dotX$.
Le equazioni descrittive di tale modello sono le seguenti:

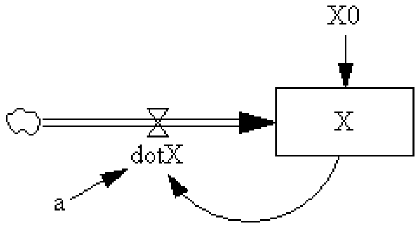

Figure 87: *Modello Vensim delle (91) (dettagli nel testo)*

```
(1) a=0.1
    Units: 1/(unit*Month) [-1,1,0.1]
(2) dotX=a*X*X
    Units: unit/Month
(3) FINAL TIME  = 10
    Units: Month
    The final time for the simulation.
(4) INITIAL TIME  = 0
    Units: Month
    The initial time for the simulation.
(5) SAVEPER  = TIME STEP
    Units: Month [0,?]
    The frequency with which output is stored.
(6) TIME STEP  = 0.125
    Units: Month [0,?]
    The time step for the simulation.
(7) X= INTEG (dotX,X0)
    Units: unit
(8) X0=1
    Units: unit
```

Le figure 88 e 89 riportano due simulazioni del modello di Figura 87, la seconda delle quali con i valori dei parametri specificati nelle equazioni precedenti, la prima con un valore negativo del parametro a.

4.5 Alcune considerazioni teoriche

Nelle sezioni precedenti abbiamo visto come ad una equazione differenziale del primo ordine possa essere associato, con pochi e semplici passaggi, un modello Vensim. Nel caso di equazioni differenziali di secondo e terzo ordine abbiamo visto come sia possibile passare agli equivalenti sistemi di due o tre equazioni differenziali del primo ordine da cui arrivare ai corrispondenti modelli Vensim. Nel caso di una equazione differenziale del terzo ordine del tipo:

$$Y''' = aY'' + bY' + cY + d \tag{92}$$

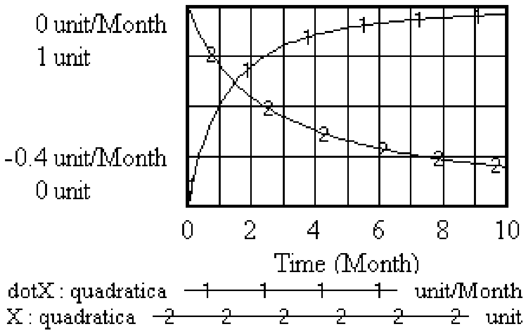

Figure 88: *Simulazione del modello di Figura 87 (dettagli nel testo)*

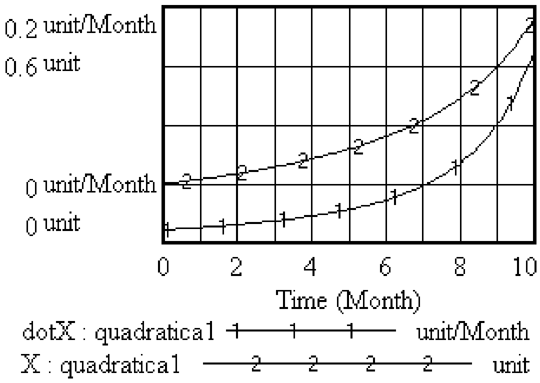

Figure 89: *Altra simulazione del modello di Figura 87 (dettagli nel testo)*

abbiamo visto che il corrispondente sistema di tre equazioni differenziali del primo ordine prevede l'uso di tre variabili di tipo **livello** fra di loro **interagenti**, nel senso che ciascuno di essi compare a secondo membro di almeno una delle equazioni differenziali del primo ordine.

$$\begin{cases} Y'(t) = X(t) \\ X'(t) = Z(t) \\ Z'(t) = aZ(t) + bX(t) + cY(t) + d \end{cases} \tag{93}$$

Se si considerano le (93) si vede facilmente come sia:

$$Z \succ X \succ Y \tag{94}$$

e:

$$X \succ Z \tag{95}$$

$$Y \succ Z \tag{96}$$

in cui la relazione $a \succ b$ significa che la variabile a è argomento di una equazione differenziale nella variabile b.
Il viceversa non è necessariamente vero nel senso che da tre equazioni differenziali del primo ordine si può passare a una equazione differenziale del terzo ordine solo se le variabili coinvolte sono fra di loro interagenti, nel senso suddetto. Nel caso delle (93), se si rappresentano le variabili come nodi di un grafo collegati da archi che implementano la relazione $\succ$ si ottiene il grafo di Figura 90. In

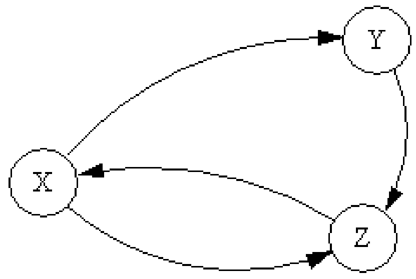

Figure 90: *Grafo che implementa la relazione $\succ$ (dettagli nel testo)*

tale grafo si ha un ciclo semplice orientato di lunghezza pari al numero dei nodi (ovvero 3) per cui ad esso si può associare una equazione differenziale del terzo ordine.
Nel caso delle seguenti equazioni differenziali:

$$\begin{cases} Y'(t) = aY(t) \\ X'(t) = bX(t) \\ Z'(t) = cZ(t) \end{cases} \tag{97}$$

ad esse corrisponde un grafo completamente sconnesso per cui ad esse non si può associare nessuna equazione differenziale di ordine superiore al primo. Nel caso delle:

$$\begin{cases} Y'(t) = aY(t) \\ X'(t) = Z(t) \\ Z'(t) = bZ(t) + cX(t) + d \end{cases} \tag{98}$$

tracciando il grafo corrispondente si individua un ciclo lungo due per cui dal sistema delle (98) si può ricavare una equazione differenziale di secondo ordine fra le variabili associate ai nodi compresi nel ciclo.

4.6 Crescita con vincoli: la curva logistica

La presente sezione introduce nei modelli limitazioni di tipo fisico alla crescita di alcune delle grandezze tipiche del modello, limitazioni che si traducono in comportamenti di tipo asintotico. La Figura 91 presenta un semplice esempio di modello in cui si hanno fenomeni di tale tipo. Le equazioni caratteristiche di tale modello sono riportate subito dopo la figura. Prima di procedere con una o più simulazioni si dà una breve descrizione sia del modello sia delle equazioni più significative.

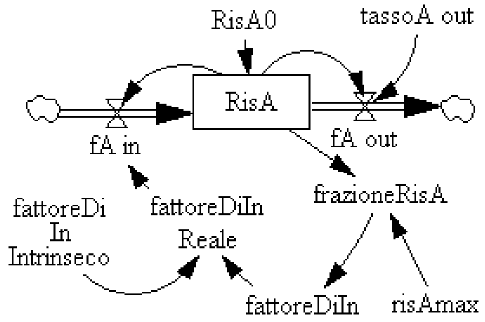

Figure 91: *Modello Vensim con saturazione (dettagli nel testo)*

```
(01)    fA in=fattoreDiInReale*RisA
        Units: ha/Year
(02)    fA out=RisA*tassoA out
        Units: ha/Year
(03)    fattoreDiIn = WITH LOOKUP (frazioneRisA,
        ([(0,0)-(1,1)],(0,1),(0.2,0.8),(0.4,0.6),
        (0.6,0.4),(0.8,0.2),(1,0) ))
        Units: Dmnl
(04)    fattoreDiInIntrinseco=1
        Units: 1/Year
(05)    fattoreDiInReale=fattoreDiIn*fattoreDiInIntrinseco
        Units: 1/Year
(06)    FINAL TIME  = 20
        Units: Year
        The final time for the simulation
(07)    frazioneRisA=RisA/risAmax
        Units: Dmnl
(08)    INITIAL TIME  = 0
        Units: Year
        The initial time for the simulation.
(09)    RisA= INTEG (fA in-fA out,RisA0)
        Units: ha [0,1000]
(10)    RisA0=10
        Units: ha
(11)    risAmax=1000
        Units: ha
(12)    SAVEPER=TIME STEP
        Units: Year [0,?]
```

```
        The frequency with which output is stored.
(13)    tassoA out=0.2
        Units: 1/Year
(14)    TIME STEP  = 1
        Units: Year [0,?]
        The time step for the simulation.
```

Il modello di Figura 91 descrive un sistema in cui una grandezza $RisA$ varia in funzione di altre due ($fAin$,$fAout$) ed è soggetta ad un valore massimo che ne limita la crescita, rappresentato dalla variabile $risAmax$. La grandezza $RisA$

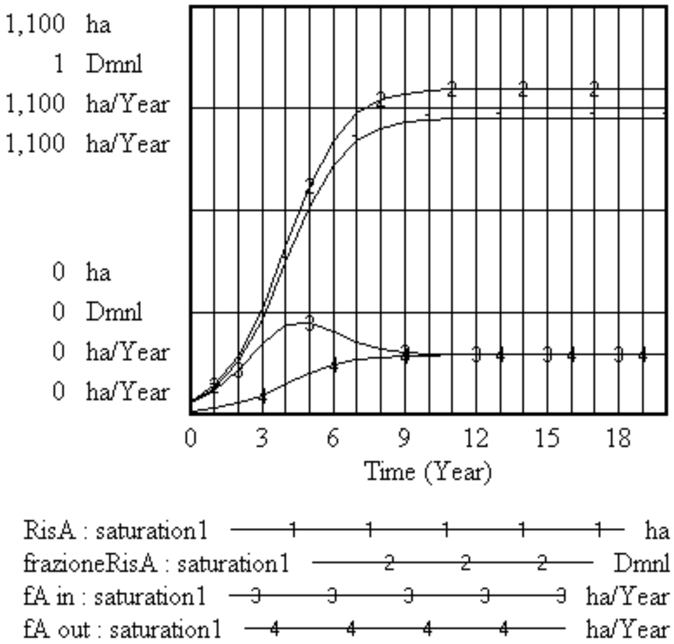

Figure 92: *Andamento di alcune grandezze del modello di Figura 91 (dettagli nel testo)*

varia in funzione di due flussi[27]:

1. fA_out che, a sua volta, dipende solo da una variabile tasso (vedi la sezione che segue) $tassoA\_out$;
2. fA_in che dipende dalle quantità:
   (a) $RisA$ ovvero quantità della risorsa correntemente disponibile;
   (b) $risAmax$ ovvero massima quantità di risorsa disponibile

[27]Come consuetudine eventuali spazi nei nomi delle variabili sono resi con il carattere _.

secondo la seguente catena di dipendenze causa÷effetto:

$$\left.\begin{array}{l} RisA \\ risAmax \end{array}\right\} \rightarrow frazioneRisA \rightarrow \left\{\begin{array}{l} fattoreDiIn \\ fattoreDiInIntrinseco \end{array}\right\} \rightarrow \% \tag{99}$$

$$\%fattoreDiInReale \rightarrow fA\_in \tag{100}$$

La variabile $frazioneRisA$ è definita dall'equazione (07) mentre la variabile $fattoreDiIn$ è definita dal'equazione (03). Si definisce un ciclo che limita la crescita di un flusso fA_in man mano che si consuma la risorsa disponibile e ci si avvicina al suo valore massimo. Se in assenza di vincoli di disponibilità si avrebbe un fattore di crescita $fattoreDiInIntrinseco$ (vedi la (04)) la presenza di un tasso di utilizzo (vedi la sezione che segue), definito dall'equazione (07), introduce un fattore di crescita reale che limita la crescita del flusso in ingresso e, quindi, la crescita della risorsa. Tali andamenti sono illustrati dalla simulazione di figura 92 in cui si vede come:

1. a fronte di un valore di $risAmax = 1000$ (vedi la (11)) la $RisA$ cresce dal valore iniziale in modo prima esponenziale per poi portarsi ad un livello di saturazione inferiore a tale valore;
2. la differenza fra $risAmax = 1000$ e $RisA$ a regime è legata alla presenza di un flusso non nullo fA_out;
3. il flusso di ingresso fA_in mostra una crescita iniziale che va oltre il valore assunto a regime in modo da consentire la rapida crescita iniziale del livello $RisA$.

Il punto chiave è che al crescere del rapporto:

$$\frac{RisA}{risAmax} \tag{101}$$

la variabile $fattoreDiIn$ tende a decrescere secondo la curva espressa dalla (03) in modo che risultino decrescere, in cascata:

1. $fattoreDiInReale$ in base alla (05),
2. fA_in in base alla (01),

in modo che la crescita di $RisA$ risulti frenata. Al momento in cui si arriva ad una condizione di equilibrio le varie grandezze $RisA$, $frazioneRisA$, fA_in, fA_out assumono valori costanti (vedi la figura 92) e li mantengono fino a che non si introducono perturbazioni nel modello agendo sulle **variabili esogene**:

1. $fattoreDiInIntrinseco$
2. $risAmax$
3. $tassoA\_out$

4.7 Alcune variabili particolari: tassi e complementi

In alcuni casi può capitare di dover usare variabili continue che assumono un valore compreso fra 0 e 1. Tali variabili sono dette **tassi**. Nelle sezioni precedenti abbiamo visto esempi di tassi sotto forma di:

1. tasso di natalità;
2. tasso di mortalità;
3. tasso di crescita.

Una variabile di tipo **tasso** ovvero un **tasso** esprime, in genere, il rapporto fra due grandezze A e B con $B \in [0, A]$ in modo che sia:

$$\frac{B}{A} \in [0, 1] \tag{102}$$

Ad esempio il tasso di natalità esprime il numero di nati su 100 individui per periodo di tempo (in genere per anno) e lo stesso vale per il tasso di mortalità. Non sempre le cose filano via così lisce e capita, a volte, di dover introdurre funzioni matematiche di massimo (max) o di minimo (min) per far si che una variabile che vogliamo definire come tasso sia forzata ad assumere un valore in $[0, 1]$.
Nelle presenti note si esaminano solo i due casi seguenti:

1. tasso di disponibilità τ_d;
2. tasso di utilizzo τ_u.

Date due grandezze x e y tali che:

1. x indica la quantità di risorsa disponibile;
2. y indica la quantità di risorsa richiesta;

sembrerebbe di poter definire:

$$\tau_d = \frac{x - y}{x} \tag{103}$$

e:

$$\tau_u = \frac{y}{x} \tag{104}$$

Così facendo non si ha la certezza che tali variabili assumano valori nell'intervallo $[0, 1]$ e, in più, si può avere che:

1. τ_d assume valori negativi e non significativi non appena si ha $x < y$;
2. τ_u assume valori maggiori di 1 e non significativi non appena si ha $x < y$.

Si noti che la condizione $x < y$ richiesta, da un punto di vista logico, per come sono state definite le due variabili, non sempre è associata ad un vincolo esplicito nel modello per cui un modo per imporla è quello di definire le variabili **tasso** in modo opportuno. Nei casi che stiamo esaminando, uno dei modi per ovviare a tali inconvenienti ed introdurre il necessario vincolo fra le variabili x e y è quello di ricorrere alle seguenti definizioni:

$$\tau_d = max\{0, \frac{x-y}{x}\} \tag{105}$$

e:

$$\tau_u = min\{1, \frac{y}{x}\} \tag{106}$$

In questo modo si evitano gli inconvenienti suddetti e si definiscono i due tassi τ_d e τ_u in modo che descrivano correttamente i fenomeni a cui si riferiscono. In questo caso fra i due tassi non si può stabilire la seguente relazione:

$$\tau_d = 1 - \tau_u \tag{107}$$

che sarebbe valida qualora non si fossero usate le funzioni *min* e *max*.
Si ritiene degno di nota segnalare come, data $x \in [0, 1]$, se si definisce:

$$y = 1 - x \tag{108}$$

si ottengono due scopi:

1. si stabilisce una relazione di proporzionalità inversa fra le due variabili (lo stesso vale anche se si introduce una costante positiva a moltiplicare il termine a secondo membro);

2. si definisce y come il complemento a 1 di x e viceversa.

La relazione:

$$y = \alpha x \tag{109}$$

(con $\alpha > 0$) si limita a definire una relazione di proporzionalità diretta fra le due variabili.

A Appendice matematica

A.1 Equazioni alle differenze e differenziali

A.1.1 Equazioni alle differenze

La base teorica delle **equazioni alle differenze** risiede nelle successioni definite per ricorrenza in cui si ha $a_{n+1} = f(a_n)$. Le equazioni alle differenze sono essenzialmente di due tipi:

1. lineari;
2. non lineari.

La forma generale è la seguente:

$$x_{n+1} = f(x_n, n) \tag{110}$$

In genere non si considera esplicitamente il parametro "tempo" **n** in modo da definire il cosiddetto **problema autonomo**:

$$x_{n+1} = f(x_n) \tag{111}$$

Nell'equazione (111) ho due soli istanti temporali che sono il "presente" e "il futuro" (il secondo dei quali dipende dal primo) per cui si parla di equazione alle differenze del primo ordine. Nel caso si abbia:

$$x_{n+1} = f(x_n, x_{n-1}, \ldots, x_{n-t}) \tag{112}$$

si parla di equazione alle differenze di ordine **t** in cui compare un istante temporale come funzione di altri t istanti immediatamente precedenti. Nel caso della (112) si ha una funzione:

$$f \,:\, \mathbb{R}^t \to \mathbb{R} \tag{113}$$

mentre nel caso della (111) si ha una funzione:

$$f \,:\, \mathbb{R} \to \mathbb{R} \tag{114}$$

A.1.2 Equazioni alle differenze: motivazioni

Vediamo di capire perchè si ricorre alle equazioni alle differenze per descrivere sistemi per i quali il tempo **t** varia "naturalmente" in modo continuo.
I motivi (alcuni, non tutti) sono essenzialmente i seguenti:

1. gusti, ovvero la preferenza di studiare un sistema utilizzando tempi discreti;
2. necessità di validare il modello utilizzando dati che sono determinati solo a scadenze precise e con una periodicità che può essere giornaliera, settimanale, mensile etc. ma che difficilmente si può pensare assimilabile ad una variazione di tipo continuo per cui nei modelli continui non si hanno dati istante per istante per la loro validazione;

3. motivi filosofici legati al fatto che si considera un sistema che varia solo a istanti discreti ed equamente distanziati.

Si ritiene degno di nota il fatto che i due modelli danno risultati diversi e che se si usa il modello continuo per ricavare degli andamenti e si operano delle discretizzazioni si ottengono risultati diversi da quelli che si ottengono partendo direttamente da un modello discreto.

A.1.3 Equazioni alle differenze lineari

Si inizia da una equazione alle differenze lineare omogenea del tipo:

$$ay_n + by_{n-1} = 0 \tag{115}$$

Se $a = 0$ e/o $b = 0$ il problema è banale. Se $a \neq 0$ posso dividere m. a m. per a in modo da scrivere:

$$y_n + cy_{n-1} = 0 \tag{116}$$

con:

$$c = \frac{b}{a} \tag{117}$$

Per la (116) si ha un valore iniziale noto:

$$y_0 = \bar{A} \tag{118}$$

Una soluzione in forma chiusa della (116) che rispetta la condizione iniziale (118) è:

$$y_n = k^n h \tag{119}$$

Si ricava la costante incognita h utilizzando il valore noto per $t = 0$ ovvero y_0 in modo che sia $h = y_0$ e quindi:

$$y_n = k^n y_0 \tag{120}$$

con $k = -c = -\frac{b}{a}$.
È facile vedere che (eseguendo quella che si chiama **analisi della stabilità**):

1. se $\mid k \mid < 1$ il sistema è asintoticamente stabile nel senso che per $n \to +\infty$ si ha $y_n \to 0$;
2. se $\mid k \mid > 1$ il sistema è instabile nel senso che per $n \to +\infty$ si ha $\mid y_n \mid \to +\infty$;
3. se $\mid k \mid = 1$ si hanno due sottocasi:
   (a) se $k = 1$ si ha $y_n = y_0$ costante, il sistema non ha evoluzione;
   (b) se $k = -1$ si ha, al variare di n, una successione di valori a segni alterni del tipo $-y_0, y_0, -y_0, y_0 \ldots$ e così via, il sistema è localmente stabile.

Più in dettaglio si ha che se $k < 0$ si hanno delle oscillazioni che:

1. se $-1 < k < 0$ sono di tipo smorzato;
2. se $k < -1$ sono crescenti in ampiezza;
3. se $k = -1$ si hanno oscillazioni di ampiezza costante.

Tale fatto viene sfruttato in economia per modellare grandezze che oscillano nel tempo come il reddito.
Nel caso si abbia:

$$ay_n + by_{n-1} = d \tag{121}$$

con d costante nota la soluzione la si determina come somma della soluzione della equazione omogenea associata e di una soluzione particolare della (121). Come visto una soluzione della equazione omogenea associata:

$$ay_n + by_{n-1} = 0 \tag{122}$$

ovvero:

$$y_n = -\frac{b}{a}y_{n-1} \tag{123}$$

che si può scrivere come:

$$y_n = ky_{n-1} \tag{124}$$

se $k = -\frac{b}{a}$ è del tipo:

$$y_n = k^n h \tag{125}$$

Per trovare una soluzione particolare della (121) si suppone $y = y_e = cost$ in modo che sia:

$$(a + b)y_e = d \tag{126}$$

da cui:

$$y_e = \frac{d}{a+b} \tag{127}$$

La soluzione generale pertanto è:

$$y_n = k^n h + \frac{d}{a+b} \tag{128}$$

A questo punto si deve calcolare h utilizzando il valore iniziale noto per $n = 0$ ovvero y_0. Con semplici calcoli si ricava:

$$h = y_0 - \frac{d}{a+b} \tag{129}$$

da cui:

$$y_n = k^n(y_0 - \frac{d}{a+b}) + \frac{d}{a+b} \tag{130}$$

Se $d = 0$ si ricade nel caso della equazione omogenea e si ha $h = y_0$.

A.1.4 Equazioni alle differenze di ordine superiore al secondo

Si passa ora a modelli più complessi cui sono associate equazioni alle differenze del tipo:

$$c_2 y_t + c_1 y_{t-1} + c_0 y_{t-2} = g(t) \tag{131}$$

in cui:

1. c_0, c_1, c_2 sono costanti date,
2. $g(t)$ è una funzione.

La soluzione generale della (131) (dal momento che l'equazione è lineare ha sempre soluzione) consiste nella somma della soluzione dell'omogenea associata (ottenuta dalla (131) ponendo $g(t) = 0$) e della soluzione particolare della (131). L'equazione omogenea associata alla (131) è la seguente:

$$c_2 y_t + c_1 y_{t-1} + c_0 y_{t-2} = 0 \tag{132}$$

che posso riscrivere come:

$$y_t + a y_{t-1} + b y_{t-2} = 0 \tag{133}$$

se pongo:

1. $c_2 \neq 0$,
2. $a = \frac{c_1}{c_2}$,
3. $b = \frac{c_0}{c_2}$.

Come soluzione della (133) si può provare con la $y_t = \lambda^t$. Sostituendo nella (133) si ha:

$$\lambda^{t-2}(\lambda^2 + a\lambda + b) = 0 \tag{134}$$

per ogni t. Supponendo $\lambda^{t-2} \neq 0$ si ha:

$$\lambda^2 + a\lambda + b = 0 \tag{135}$$

ovvero:

$$\lambda_1, \lambda_2 = \frac{-a \pm \sqrt{\Delta}}{2} \tag{136}$$

con:

$$\Delta = a^2 - 4b \tag{137}$$

Si hanno i casi seguenti:

1. $\Delta > 0$ le due radici sono reali e distinte;
2. $\Delta = 0$ le due radici sono coincidenti;
3. $\Delta < 0$ le due radici sono complesse coniugate.

Nel primo caso la soluzione dell'omogenea è del tipo:

$$y_t = A_1\lambda_1^t + A_2\lambda_2^t \tag{138}$$

in cui (essendo l'equazione del secondo ordine) si ha che i parametri A_1 e A_2 vanno determinati in base a due condizioni iniziali.
Nel secondo caso le due radici sono coincidenti per cui la soluzione dell'omogenea è del tipo:

$$y_t = A_1\lambda_1^t + A_2 t\lambda_1^t \tag{139}$$

ciò perchè la soluzione deve essere combinazione lineare degli elementi di una base che, in questo caso, sono λ_1^t e $t\lambda_1^t$.
Nel terzo caso le radici sono complesse coniugate per cui la soluzione dell'omogenea è del tipo:

$$y_t = r^t(A_1 cos\omega t + A_2 sen\omega t) \tag{140}$$

in cui r rappresenta il modulo delle soluzioni e ω l'argomento. Più in dettaglio si ha che le soluzioni sono esprimibili come $a \pm i\theta$ in modo che sia:

1. $r^2 = a^2 + \theta^2$,
2. $rcos\omega = a$,
3. $rsen\omega = \theta$.

Dalle ultime due relazioni si ha:

$$\omega = arctg(\frac{\theta}{a}) \tag{141}$$

Alla soluzione dell'omogenea va sommata la soluzione particolare della:

$$c_2 y_t + c_1 y_{t-1} + c_0 y_{t-2} = F \tag{142}$$

con F costante. La soluzione la si trova imponendo:

$$y_t = y_{t-1} = y_{t-2} = y^* \tag{143}$$

per ogni t in modo da ricavare:

$$y^* = \frac{F}{c_2 + c_1 + c_0} \tag{144}$$

La condizione di stabilità si traduce nelle due seguenti:

1. $\mid \lambda_1 \mid < 1$,
2. $\mid \lambda_2 \mid < 1$.

Si fa notare che:

1. al crescere dell'ordine di una equazione alle differenze cresce il grado della corrispondente equazione caratteristica;
2. in economia dare valori ai parametri a e b presenta dei problemi;
3. si hanno condizioni necessarie e sufficienti per la stabilità delle soluzioni di una equazione alle differenze di ordine n senza che sia necessario calcolare effettivamente tali soluzioni;
4. nelle equazioni lineari esiste sempre la soluzione stazionaria;
5. si può sempre passare da una equazione alle differenze di ordine n ad un sistema di n equazioni alle differenze del primo ordine.

In merito all'ultimo punto si prende in esame il caso di $n = 2$. Data una equazione alle differenze del secondo ordine del tipo:

$$c_2 y_t + c_1 y_{t-1} + c_0 y_{t-2} = g(t) \tag{145}$$

la si può trasformare nel seguente sistema di due equazioni del primo ordine:

$$\begin{cases} x_t = y_{t-1} \\ c_2 y_t + c_1 x_t + c_0 x_{t-1} = g(t) \end{cases} \tag{146}$$

Per equazioni alle differenze di grado superiore al secondo si ha il criterio di stabilità noto come **Routh Hurwitz stability condition**.

A.1.5 Equazioni differenziali lineari a variabili separabili

Una equazione si dice differenziale se contiene le derivate di diversi ordini di una variabile rispetto al tempo t ovvero se ha la forma seguente:

$$f(y', y", y''', \ldots, y^n, y, t) = 0 \tag{147}$$

Nei problemi di tipo economico la (147) assume di solito la forma seguente:

$$f(y', y", y''', \ldots, y^n, y) = 0 \tag{148}$$

in modo da descrivere un problema detto **autonomo** in quanto non dipendente in modo esplicito dal parametro tempo t. La (147) può essere espressa in forma esplicita come:

$$y^n = h(y', y", y''', \ldots, y^{n-1}, y) \tag{149}$$

una cui forma ulteriormente semplificata è la seguente:

$$y' = g(t) \tag{150}$$

con $y = y(t)$ e $n = 1$. Semplificando ulteriormente si arriva alle equazioni differenziali lineari del tipo:

$$ay' + by = g(t) \tag{151}$$

La (151) non è una forma autonoma dal momento che vi compare $g(t)$ e la sua soluzione è data dalla somma della soluzione dell'equazione omogenea associata e della soluzione particolare della (151).
Vediamo la soluzione dell'omogenea associata:

$$ay' + by = 0 \tag{152}$$

Suppongo $a \neq 0$ e separo le variabili in modo da ottenere:

$$\frac{dy}{y} = -cdt \tag{153}$$

con:

$$c = \frac{b}{a} \tag{154}$$

e $y \neq 0$. La soluzione $y = 0$ (ammissibile per la (152)) viene detta **soluzione singolare**. La soluzione della (153) è:

$$ln \mid y \mid = -ct + k \tag{155}$$

ovvero:

$$\mid y \mid = k' e^{-ct} \tag{156}$$

dove $k' = e^k$. Per ricavare il valore di k' si ricorre al valore iniziale y_0.
Si ha (per semplicità si suppone $y_0 = y(0)$):

$$k' = \mid y_0 \mid \tag{157}$$

da cui, inserendo tale valore nella (156):

$$y = y_0 e^{-ct} \tag{158}$$

Si nota che (nell'ipotesi che sia $c > 0$):

1. se $y_0 > 0$ la (158) assume valori positivi per $t \geq 0$ e tendenti a 0 per $t \longrightarrow +\infty$;

2. se $y_0 < 0$ la (158) assume valori negativi per $t \geq 0$ e tendenti a 0 per $t \longrightarrow +\infty$.

La stabilità della soluzione (che non mostra mai un comportamento oscillante) dipende anche dal segno di c: se $c < 0$ si hanno infatti comportamenti divergenti in cui la y in modulo cresce in modo illimitato.
Si trova la soluzione particolare supponendo $g(t) = A$ costante in modo da risolvere la:

$$ay' + by = A \tag{159}$$

con $y = cost$ ovvero $y' = 0$ e quindi:

$$y = \frac{A}{b} \tag{160}$$

La soluzione generale della (151) è quindi:

$$y = he^{-ct} + \frac{A}{B} \tag{161}$$

Con la condizione iniziale per $t = 0$ si ricava h ovvero:

$$h = y_0 - \frac{A}{B} \tag{162}$$

e infine:

$$y = (y_0 - \frac{A}{B})e^{-ct} + \frac{A}{B} \tag{163}$$

Di nuovo, se $A = 0$ siamo nel caso di una equazione omogenea e quindi $h = y_0$.

A.1.6 Equazioni differenziali del secondo ordine (cenni)

Una equazione differenziale lineare del secondo ordine ha la forma:

$$a_2\frac{d^2y}{dt^2} + a_1\frac{dy}{dt} + a_0y = g(t) \tag{164}$$

cui corrisponde la seguente equazione caratteristica:

$$a_2\lambda^2 + a_1\lambda + a_0 = 0 \tag{165}$$

con radici:

$$\lambda_{1,2} = \frac{-a_1 \pm \sqrt{\Delta}}{2a_2} \tag{166}$$

con:

$$\Delta = a_1^2 - 4a_2a_0 \tag{167}$$

Si hanno i seguenti casi:

1. $\Delta > 0$ le due radici sono reali e distinte per cui si hanno gli esponenziali $e^{\lambda_1 t}$ e $e^{\lambda_2 t}$ con andamenti crescenti o decrescenti in funzione dei segni delle radici;

2. $\Delta = 0$ le due radici sono reali e coincidenti (ovvero $\lambda_1 = \lambda_2 = \lambda$) per cui si hanno gli esponenziali $e^{\lambda t}$ e $te^{\lambda t}$ (in modo da avere anche in questo caso due elementi linearmente indipendenti) con andamenti crescenti o decrescenti in funzione del segno della radice;

3. $\Delta < 0$ le due radici sono complesse coniugate $a \pm bi$ per cui si hanno soluzioni oscillanti con andamenti crescenti o decrescenti in funzione del segno della parte reale a.

La soluzione della (164) è data dalla somma della soluzione dell'equazione omogenea associata ovvero della:

$$a_2\frac{d^2y}{dt^2} + a_1\frac{dy}{dt} + a_0y = 0 \tag{168}$$

e una soluzione particolare della (164) ottenuta imponendo $g(t) = F = costante$ e $y_e = costante$ in modo che le derivate si annullino e si abbia (con semplici sostituzioni e pochi calcoli):

$$y_e = \frac{F}{a_0} \tag{169}$$

A.1.7 Equazioni alle differenze

Si danno alcuni esempi di equazioni alle differenze con applicazioni all'economia per poi passare ad alcune questioni di teoria ([Cam04]).

Esempio A.1 *Se, dato un conto corrente, $x(t)$ è il capitale all'istante t ad un tasso di interesse costante i si vuole il capitale all'istante $t+1$. È facile vedere come (nell'ipotesi che il capitale iniziale in $t = 0$ sia $x(0) = C_0$ e che si ridepositino sul conto corrente in ogni periodo gli interessi del periodo precedente) sia:*

$$x(t+1) = (1+i)x(t) \tag{170}$$

Usando una notazione ormai familiare si ha:

$$x_{t+1} = (1+i)x_t \tag{171}$$

con $x_0 = C_0$.
Si ha una equazione alle differenze:

1. *del primo ordine perchè l'evoluzione dipende da un solo istante di tempo ovvero t;*
2. *lineare perchè x_{t+1} e x_t hanno un esponente pari a 1;*
3. *a coefficienti costanti se i è costante.*

Una soluzione della (170) è la seguente:

$$x_t = (1+i)^t x_0 \tag{172}$$

valida per $t = 0, 1, 2, \ldots$.

Esempio A.2 *Se x_t è la grandezza di una popolazione all'istante t e k ne rappresenta il tasso di crescita sotto l'ipotesi che la popolazione non abbia ulteriori vincoli si ha che la grandezza della popolazione all'istante $t+1$ è:*

$$x_{t+1} = kx_t \tag{173}$$

con un valore iniziale $x_0 = P$. La (173) ha la seguente soluzione:

$$x_t = k^t P \tag{174}$$

Si hanno i casi seguenti:

1. *$k > 0$ la popolazione cresce in modo esponenziale,*

2. $k < 0$ *la popolazione decresce in modo esponenziale.*

Esempio A.3 *Se si ha un debito* D_t *che cresce ad un tasso di interesse costante e pari a* r *si ha:*

$$D_{t+1} = D_t + rD_t = (1 + r)D_t \tag{175}$$

la cui soluzione, dato un valore iniziale per $t = 0$ D_0 *è simile a quella già vista per il primo esempio di questa sezione.*

In ogni caso, data una equazione alle differenze, ciò che si vuole fare è:

1. calcolare, se possibile, una soluzione in forma chiusa dell'equazione in modo che sia possibile calcolare il valore della variabile in ogni istante noto il valore iniziale e i valori dei parametri che compaiono nell'equazione;

2. valutare la stabilità della soluzione trovata.

In merito alle equazioni alle differenze si ha:

1. l'ordine di una equazione è dato dalla differenza fra il pedice di valore più alto della variabile e quello di valore più basso, negli esempi visti in questa sezione tale valore è sempre stato pari ad 1 per cui le equazioni viste sono del primo ordine;

2. data una equazione alle differenze si può sempre operare una traslazione dei tempi in modo che, in una equazione del primo ordine, sia sempre possibile passare dai pedici t e $t - 1$ ai pedici $t + 1$ e t;

3. una equazione alle differenze di ordine n per essere risolta abbisogna di n condizioni iniziali;

4. una equazioni alle differenze di ordine n ha infinite soluzioni, una per ogni possibile $n-$upla di condizioni iniziali, date le n condizioni iniziali una soluzione è rappresentata da una successione di numeri;

5. la stabilità di una soluzione dipende in genere dai valori dei coefficienti e non dai valori iniziali.

Per prima cosa, quindi, si vuole una regola generale che esprima x_t in funzione di t, dei parametri dell'equazione e delle condizioni iniziali, ovvero si vuole una **soluzione in forma chiusa**. Nel seguito ci riferiremo quasi esclusivamente alle equazioni alle differenze del primo ordine. Vediamo per prima cosa due esempi.

Esempio A.4 *Se si ha:*

$$x_{t+1} = ax_t \tag{176}$$

per $t = 0, 1, 2, \ldots,$ *con un valore iniziale* x_0 *si vede facilmente (per induzione) che la soluzione in forma chiusa della (176) è:*

$$x_t = a^t x_0 \tag{177}$$

Si ha:

1. per ogni x_0 si ha una soluzione (condizione ai valori iniziali);

2. se $| a |< 1$ per $t \longrightarrow \infty$ $x_t \longrightarrow 0$

Esempio A.5 *Se si ha:*

$$x_{t+1} = \frac{x_t}{1 + x_t} \tag{178}$$

per $t = 0, 1, 2, \ldots,$ con un valore iniziale x_0 si vede facilmente (per induzione) che la soluzione in forma chiusa della (176) è:

$$x_t = \frac{x_0}{1 + tx_0} \tag{179}$$

In questo caso per $t \longrightarrow \infty$ $x_t \longrightarrow 0$ per ogni valore iniziale non nullo x_0.

Una equazione alle differenze del primo ordine ha la seguente forma generale:

$$x_{t+1} = f(t, x_t) \tag{180}$$

con $t \in \mathbb{N}$[28], un valore iniziale x_0 e con la funzione:

$$f \, : \, \mathbb{N} \times \mathbb{R} \longrightarrow \mathbb{R} \tag{181}$$

La presenta di t nella (181) classifica l'equazione come non autonoma. La soluzione di una equazione alle differenze finite è una successione di valori che inizia per x_0 e soddisfa la (181) che ha una soluzione (ovvero descrive una successione diversa di valori) per ogni x_0 fissato.
Una equazione alle differenze finite si dice.

1. lineare del primo ordine se è del tipo:

$$x_{t+1} = a(t)x_t + b(t) \tag{182}$$

   con:

$$a \, : \, \mathbb{R} \longrightarrow \mathbb{R} \tag{183}$$

$$b \, : \, \mathbb{R} \longrightarrow \mathbb{R} \tag{184}$$

2. lineare del primo ordine omogenea se è del tipo:

$$x_{t+1} = a(t)x_t \tag{185}$$

Vediamo il caso di una equazione omogenea. Si ha:

$$x_{t+1} = a(t)x_t \tag{186}$$

In questo caso lo spazio delle soluzioni x_t è uno spazio vettoriale di dimensione 1 ovvero le soluzioni della (186) sono fra loro proporzionali per cui:

[28] per semplicità di notazione si suppone che $\mathbb{N}$ contenga lo 0.

1. la successione nulla è una soluzione;
2. se y_t è una soluzione lo è anche ky_t per ogni $k \in \mathbb{R}$;
3. se y_t e z_t sono soluzioni allora esiste $k \in \mathbb{R}$ tale che $y_t = kz_t$.

Nel caso di una equazione non omogenea:

$$x_{t+1} = a(t)x_t + b(t) \tag{187}$$

ad essa si associa un'equazione omogenea:

$$x_{t+1} = a(t)x_t \tag{188}$$

in modo che le soluzioni della (187) siano del tipo:

$$x_t = y_t + z_t \tag{189}$$

in cui:

1. y_t è una soluzione dell'equazione omogenea associata (188);
2. z_t è una soluzione particolare della (187).

Si ha una ulteriore semplificazione se $a(t) = a$ e $b(t) = b$. In questo caso si parla di equazioni alle differenze finite a coefficienti costanti che assumono la forma:

$$x_{t+1} = ax_t + b \tag{190}$$

con un valore iniziale x_0. La soluzione generale della (190) è data dalla somma di una soluzione dell'equazione omogenea associata:

$$x_{t+1} = ax_t \tag{191}$$

e di una soluzione particolare della (190). Nel nostro caso si ha che la soluzione della (191) è:

$$x_t = ha^t \tag{192}$$

mentre una soluzione particolare della (190) è:

$$x^* = \frac{b}{1-a} \tag{193}$$

in modo che la soluzione generale sia:

$$x_t = ha^t + \frac{b}{1-a} \tag{194}$$

Se si sfrutta la condizione iniziale per $t = 0$ x_0 si può ricavare h come:

$$h = x_0 - \frac{b}{1-a} \tag{195}$$

in modo che la soluzione cercata sia:

$$x_t = (x_0 - \frac{b}{1-a})a^t + \frac{b}{1-a} \tag{196}$$

Di nuovo si noti che se $b = 0$ tutto si riduce alla soluzione di una equazione alle differenze finite omogenea per cui la soluzione è:

$$x_t = ha^t \tag{197}$$

con $h = x_0$. Quanto visto vale nel caso che sia $a \neq 1$. Nel caso sia $a = 1$ l'equazione risulta:

$$x_{t+1} = x_t + b \tag{198}$$

per cui una soluzione particolare la si cerca fra quelle del tipo $x^* = kt$ ovvero.

$$k(t+1) = kt * b \tag{199}$$

da cui $k = b$ per cui $x^* = bt$. In merito alla soluzione dell'omogenea si ha che vale x_0 per cui la soluzione generale della (198) è:

$$x_t = x_0 + bt \tag{200}$$

Un caso leggermente più complesso è il seguente:

$$x_{t+1} = ax_t + b(t) \tag{201}$$

con una condizione iniziale nota x_0. Per la soluzione dell'omogenea associata non ci sono variazioni. Si deve considerare, invece, come sia possibile ricavare una soluzione particolare della (201). In queste note si esaminano solo due casi:

1. $b(t)$ è un polinomio di grado n anche non completo;
2. $b(t) = \beta\alpha^t$.

Nel primo caso si hanno due sottocasi:

1. se $a \neq 1$ si cerca una soluzione particolare del tipo $x^* = P(t)$ con $P(t)$ polinomio completo di grado n;
2. se $a = 1$ si cerca una soluzione particolare del tipo $x^* = tP(t)$ con $P(t)$ polinomio completo di grado n;

Una volta sostituita l'espressione della x^* nella (201) e svolti i calcoli si applica il principio di eguaglianza fra polinomi (in base al quale i coefficienti dei termini di grado uguale devono essere uguali) per ricavare i coefficienti incogniti delle potenze di $P(t)$.
Anche nel secondo caso si hanno due sottocasi:

1. se $a \neq \alpha$ la soluzione particolare della (201) è del tipo $x^* = k\alpha^t$;
2. se $a0\alpha$ la soluzione particolare della (201) è del tipo $x^* = kt\alpha^t$.

Anche in questo caso si ricava k applicando il principio di eguaglianza fra polinomi.
A questo punto diciamo qualcosa in merito al concetto di **punto di equilibrio** di una equazione alle differenze finite. Un punto di equilibrio è, per definizione, una soluzione costante dell'equazione. Se ad esempio ho:

$$x_{t+1} = ax_t + b \tag{202}$$

e $a \neq 1$ ho che $x^* = k$ è punto di equilibrio della (202) sse:

$$x^* = \frac{b}{1-a} \tag{203}$$

In più si ha che tale soluzione è **globalmente asintoticamente stabile** se:

$$lim_{t \longrightarrow +\infty} x_t = x^* \tag{204}$$

mentre è instabile in ogni altro caso.

Esempio A.6 *Supponiamo di avere:*

$$x_{t+1} = ax_t + b \tag{205}$$

con $a \neq 1$ e una condizione iniziale x_0 nota. Come punto di equilibrio si ha:

$$x^* = \frac{b}{1-a} \tag{206}$$

mentre la soluzione generale della (205) è:

$$x_t = a^t h + \frac{b}{1-a} \tag{207}$$

Ricavando h dalla condizione iniziale si ha:

$$h = x_0 - \frac{b}{1-a} \tag{208}$$

per cui la soluzione generale ha la forma:

$$x_t = a^t(x_0 - \frac{b}{1-a}) + \frac{b}{1-a} \tag{209}$$

Se $\mid a \mid < 1$ si ha:

$$lim_{t \longrightarrow +\infty} x_t = \frac{b}{1-a} \tag{210}$$

per cui in tal caso il punto di equilibrio è asintoticamente (ovvero come punto limite dell'evoluzione) globalmente (ovvero per ogni valore della condizione iniziale x_0) stabile. Se $\mid a \mid > 1$ si ha:

$$lim_{t \longrightarrow +\infty} \mid x_t \mid = +\infty \tag{211}$$

per cui il punto di equilibrio è instabile. Si hanno altri due casi:

1. $a = 1$,

2. $a = -1$.

Nel primo caso l'equazione data assume la forma $x_{t+1} = x_t + b$. *La soluzione dell'omogenea è* $x_t = h$ *mentre la soluzione particolare ha la forma* $x^* = kt$. *Sostituendo si ricava* $x^* = bt$ *in modo che la soluzione generale sia:*

$$x_t = h + bt \tag{212}$$

Con la condizione iniziale si ricava h *ottenendo:*

$$x_t = x_0 + bt \tag{213}$$

Nel secondo caso l'equazione alle differenze finite data assume la forma:

$$x_{t+1} = -x_t + b \tag{214}$$

La soluzione dell'omogenea associata alla (214) è:

$$x_t = h(-1)^t \tag{215}$$

mentre la soluzione particolare è:

$$x^* = \frac{b}{2} \tag{216}$$

in modo che la soluzione generale sia:

$$x_t = h(-1)^t + \frac{b}{2} \tag{217}$$

Utilizzando la condizione iniziale si ha:

$$h = x_0 - \frac{b}{2} \tag{218}$$

da cui:

$$x_t = (x_0 - \frac{b}{2})(-1)^t + \frac{b}{2} \tag{219}$$

ovvero si hanno oscillazioni di ampiezza costante attorno al punto di equilibrio.

A.1.8 Modello della ragnatela in tempo discreto

Esempio A.7 (Modello della ragnatela in tempo discreto ([Cam04])) *Si considerano la domanda* D_t *e l'offerta* S_t *di un singolo bene sul mercato all'istante* t. *Se i parametri* a, b, c, d *sono tutti strettamente positivi si ha:*

$$D_t = b - ap_t \tag{220}$$

$$S_t = cp_{t-1} - d \tag{221}$$

*Il prezzo p_t viene fissato in modo che ci sia un esaurimento dell'offerta ovvero uno svuotamento del mercato (detto **market clearing**) ovvero in modo che sia:*

$$D_t = S_t \tag{222}$$

ovvero (con uno shift in avanti di un periodo):

$$ap_{t+1} + cp_t = b + d \tag{223}$$

da cui :

$$p_{t+1} = -\frac{c}{a}p_t + \frac{b+d}{a} \tag{224}$$

La soluzione generale della (224) è data dalla somma di una soluzione particolare della (224) e della soluzione dell'equazione omogenea associata. Si ha quindi:

$$p_t = h(-\frac{c}{c})^t + \frac{b+d}{a+c} \tag{225}$$

Al solito si ricava h utilizzando la condizione iniziale per $t = 0$ p_0 in modo che sia:

$$h = p_0 - \frac{b+d}{a+c} \tag{226}$$

e quindi:

$$p_t = (p_0 - \frac{b+d}{a+c})(-\frac{c}{a})^t + \frac{b+d}{a+c} \tag{227}$$

A questo punto si vuole sapere se esiste un prezzo p^* a partire dal quale il bene viene scambiato ad un prezzo costante in modo che si abbia sempre una uguaglianza di offerta e domanda. Tale prezzo è dato dalla soluzione particolare della (224) per cui si ha:

$$p^* = \frac{b+d}{a+c} \tag{228}$$

A questo punto si vuole sapere se a partire da p_0 la successione dei prezzi p_t converge a no a p^*. In pratica ci si chiede se esiste un prezzo di equilibrio (e la (228) ne è la risposta) e se tale prezzo è stabile o meno. Dal momento che si ha $c > 0$ e $a > 0$ si ha che p_t converge a p^* se e solo se:

$$\frac{c}{a} < 1 \tag{229}$$

Se, invece, $c > a$ in modo che:

$$\frac{c}{a} > 1 \tag{230}$$

in base alla:

$$p_t = (p_0 - \frac{b+d}{a+c})(-\frac{c}{a})^t + \frac{b+d}{a+c} \tag{231}$$

si hanno oscillazioni di ampiezza crescente attorno al prezzo di equilibrio. Se, infine, $c = a$ si ha:

$$p_t = (p_0 - p^*)(-1)^t + p^* \tag{232}$$

in modo che:

1. per $t = 2n$ si ha $p_t = p_0$,

2. per $t = 2n + 1$ si ha $p_t = -p_0 + 2p^*$, dove $n \in \mathbb{N}$.

Esempio A.8 (Montante a capitalizzazione composta) *Si deposita un capitale C ad un tasso di interesse costante i. Se M_t è il montante all'istante t e M_{t+1} è il montante all'istante $t + 1$ si ha:*

$$M_{t+1} = M_t + iM_t = (1 + i)M_t \tag{233}$$

con $M_0 = C$. La soluzione generale è:

$$M_t = h(1 + i)^t \tag{234}$$

Dalla condizione iniziale si ricava subito che $h = C$.

Esempio A.9 (Montante di una rendita periodica posticipata) *Si versano n rate di importo R fisso con un tasso di interesse i costante. Si vuole il montante all'istante t. Si ha:*

$$M_{t+1} = M_t + iM_t + R = M_t(1 + i) + R \tag{235}$$

La soluzione della (235) è data da una soluzione particolare più la soluzione dell'equazione omogenea associata. La soluzione dell'omogenea è $M_t = (1+i)^t h$. Una soluzione particolare è:

$$M = M(1 + i) + R \tag{236}$$

ovvero:

$$M = -\frac{R}{i} \tag{237}$$

per cui la soluzione generale è:

$$M_t = (1 + i)^t h - \frac{R}{i} \tag{238}$$

ricavando h dalla condizione iniziale si ha:

$$h = \frac{R}{i} \tag{239}$$

e quindi:

$$M_t = (1 + i)^t \frac{R}{i} - \frac{R}{i} = \frac{R}{i}[(1 + i)^t - 1] \tag{240}$$

A.1.9 Sistemi di equazioni lineari alle differenze

Si parte dalla seguente equazione:

$$X_{t+1} = AX_t + b \tag{241}$$

in cui X_{t+1}, X_t e b sono vettori di tipo $n \times 1$ e A è una matrice di tipo $n \times n$. Trovare una soluzione della (241) significa trovare una soluzione in forma chiusa del tipo:

$$X_t = f(t) \tag{242}$$

Se si vuole la soluzione particolare si impone che $X \in \mathbb{R}^n$ sia costante ovvero che sia:

$$X_t = X \tag{243}$$

per ogni t. La soluzione costante X deve verificare la seguente equazione:

$$X = AX + b \tag{244}$$

ovvero:

$$(I - A)X = b \tag{245}$$

per cui, sotto l'ipotesi che la $I - A$ sia invertibile, si ottiene:

$$X = (I - A)^{-1}b \tag{246}$$

Al proposito si ha il seguente teorema di esistenza e unicità:

Theorema A.1 *Se la $I-A$ è invertibile allora la (244) ha soluzione (data dalla (246)) e tale soluzione è unica.*

Supponiamo di avere un problema di tipo omogeneo (in cui cioè si ha $b = 0$) ovvero del tipo:

$$X_{t+1} = AX_t \tag{247}$$

In questo caso si ha il seguente Teorema.

Theorema A.2 *Se nella (247) la matrice $A \in \mathbb{R}^{n\times n}$ è diagonalizzabile si hanno n autovalori λ_i e n autovettori v_i ($i = 1, \ldots, n$) linearmente indipendenti in base ai quali si ottiene la soluzione della (247) come:*

$$X_t = \sum_{i=1}^{n} c_i \lambda_i^t v_i \tag{248}$$

con $c_i \in \mathbb{R}$.

Se non si hanno le condizioni iniziali nella soluzione compaiono i parametri c_i mentre, in presenza di n condizioni iniziali, le si usa per esplicitare i c_i. Nel caso si abbia:

$$X_{t+1} = AX_t + b \tag{249}$$

la soluzione è data dalla somma della soluzione dell'equazione omogenea associata (247) e della soluzione particolare (246) per cui si ha:

$$X_t = \sum_{i=1}^{n} c_i \lambda_i^t v_i + (I - A)^{-1}b \tag{250}$$

A.1.10 Analisi della stabilità

Le condizioni sulla stabilità globale della (250) sono relative agli autovalori λ_i. Si hanno al proposito le seguenti condizioni.

1. Se $\forall i \mid \lambda_i \mid < 1$ si parla di pozzo in quanto tutte le traiettorie convergono alla soluzione particolare (246). In questo caso i λ_i sono attrattori. Si parla, in questo caso, di **stabilità asintotica** dal momento che la soluzione arriva alla $(I - A)^{-1}$ solo all'infinito. La $(I - A)^{-1}$ viene detta **soluzione stazionaria**.

2. Se $\forall i \mid \lambda_i \mid > 1$ le traiettorie divergono dalla soluzione particolare e i λ_i sono detti repulsori.

3. Se esiste i tale che $\lambda_i = 1$ si hanno problemi (vedi oltre).

4. Se la condizione è che alcuni λ_i hanno modulo minore di 1 mentre altri hanno modulo maggiore di 1 ci si trova in una condizione di sella.

Vediamo più in dettaglio l'ultima condizione.
Suppongo, dopo una eventuale rimunerazione degli autovalori, che sia:

1. per $i = 1, \ldots, k \mid \lambda_i \mid < 1$,

2. per $i = k + 1, \ldots, n \mid \lambda_i \mid > 1$.

Sotto tali ipotesi si vuole sapere come si comportano le traiettorie. Si può vedere che si allontanano dalla soluzione di equilibrio per cui si ha instabilità però la situazione merita di essere analizzata più in dettaglio. Per capire come vanno le cose è necessario avere presente che gli autovettori sono **linearmente indipendenti** per cui esistono traiettorie particolari che si avvicinano alla soluzione stabile ed altre che se ne allontanano. Se ho k autovalori con modulo minore di uno ho che esiste una varietà stabile di dimensione k che tende alla soluzione stabile:

$$X^* = (I - A)^{-1} b \tag{251}$$

Vediamo un esempio in $\mathbb{R}^2$ con due autovalori:

1. $\mid \lambda_1 \mid < 1$,

2. $\mid \lambda_2 \mid > 1$

cui corrispondono gli autovettori v_1 e v_2 che sono una base di $\mathbb{R}^2$.
La figura 93 illustra un esempio di tale situazione. Le traiettorie del tipo (1) sono quelle su cui ha effetto solo $\mid \lambda_1 \mid < 1$ (ovvero sono una varietà stabile di dimensione 1 su cui considero solo l'effetto dell'autovettore v_1) mentre quelle di tipo (2) sono quelle su cui ha effetto solo $\mid \lambda_2 \mid > 1$ (ovvero sono una varietà instabile di dimensione 1 su cui considero solo l'effetto dell'autovettore v_2). Componendo due traiettorie, una per ciascun tipo, si ottengono le traiettorie "divergenti" di figura che si allontano dal punto di sella X^* secondo una traiettoria del tipo $xy = k$. Un tale comportamento in fisica è detto **sella instabile**

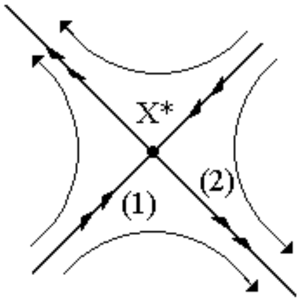

Figure 93: *Traiettorie stabili e instabili con punto di sella (dettagli nel testo)*

mentre in economia è detto semplicemente **sella** e lo si vede, a seconda dei casi, sia come sinonimo di stabilità sia come sinonimo di instabilità dal momento che, in economia, gli agenti (supposti intelligenti e autonomi) sanno come porsi su una varietà stabile come (1) o su una instabile come (2)).
Per completare l'analisi, a rigore, andrebbero esaminati i vari sottocasi:

1. autovalori reali;
2. autovalori complessi coniugati con parte reale positiva, nulla o negativa;

e così via.
Nel caso si abbia una versione bidimensionale (ovvero in $\mathbb{R}^2$) ad essa corrisponde una sola equazione alle differenze del primo ordine del tipo:

$$x_{t+1} = ax_t + b \tag{252}$$

La soluzione stazionaria della (252), con tecniche standard, la si trova supponendo $x_t = x$ per cui è la soluzione di:

$$x = ax + b \tag{253}$$

ovvero:

$$x = \frac{b}{1-a} \tag{254}$$

La soluzione generale della (252) la si trova come soluzione iterata del seguente sistema equivalente di equazioni:

$$\begin{cases} x_{t+1} = x_t \\ x_{t+1} = ax_t + b \end{cases} \tag{255}$$

La prima deriva dall'imporre la condizione $x = costante$ per ogni t.
La figura 94 illustra alcuni passi del procedimento iterativo rappresentato dalle equazioni (255): la prima equazione è denotata con (1) in figura e la seconda

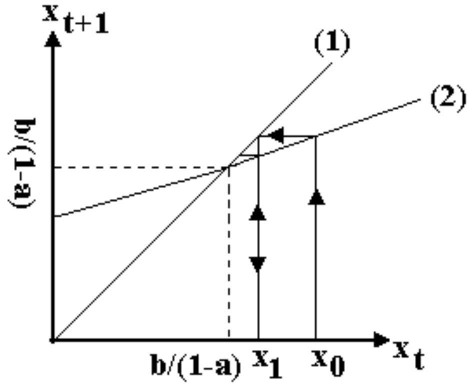

Figure 94: *Procedimento iterativo, caso stabile (dettagli nel testo)*

con (2) mentre la soluzione stazionaria (253) stà sull'incrocio fra le due rette. Partendo da un valore iniziale x_0 si calcola x_1 mediante la prima equazione e, usando tale valore, si calcola x_2 e così via. Nel caso di figura si vede che la successione di punti converge alla soluzione stazionaria per cui il punto è stabile. La condizione di stabilità è che la retta (2) abbia una pendenza minore di 1 ovvero che sia $a < 1$: infatti la soluzione dell'equazione omogenea associata è ca^t che tende a 0 per $t \longrightarrow +\infty$ solo se $a < 1$. Se $a > 1$ si ha la situazione di figura 95 in cui la traiettoria è divergente nel senso che i punti, a partire dal valore iniziale x_0, si allontanano sempre più dalla soluzione stazionaria $x^* = \frac{b}{1-a}$.

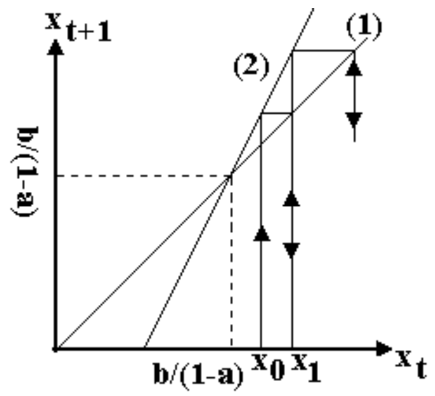

Figure 95: *Procedimento iterativo, caso instabile (dettagli nel testo)*

A.1.11 Sistemi di equazioni non lineari

Cosa accade nel caso si abbia un sistema di equazioni non lineari del tipo:

$$X_{t+1} = f(X_t) \tag{256}$$

in cui:

1. X_t e X_{t+1} sono elementi di $\mathbb{R}^n$,
2. la f è una funzione non lineare?

Si ha il Teorema di Hartman-Grobman che, in modo informale, stabilisce quanto segue: se si escludono i casi in cui si hanno autovalori di modulo uguale a 1 (ovvero i casi in cui $\exists i$ tale che $\mid \lambda_i \mid= 1$) si ha che il **comportamento qualitativo** delle orbite descritte dal sistema non lineare coincide con quello delle orbite del sistema linearizzato. Tale teorema si basa sul presupposto della linearizzazione del sistema (256) e sull'ipotesi che ci si accontenti di informazioni relative ad un intorno del punto di equilibrio $X = f(X)$ in modo che si possa affermare che il comportamento del sistema sia equivalente a quello descritto dalla sua forma linearizzata. La condizione:

$$\mid \lambda_i \mid< 1 \tag{257}$$

per ogni $i = 1, \dots, n$, si traduce nel fatto che il punto di equilibrio è stazionario. Anche nel caso di sistemi non lineari si vogliono i punti di equilibrio stabili e, a tal fine, si sfruttano informazioni locali in un intorno di ogni punto di equilibrio associate ad una linearizzazione del sistema.
Vediamo un esempio in $\mathbb{R}^2$ (vedi la figura 96).
Si abbia:

$$x_{t+1} = x_t^2 \tag{258}$$

Le soluzioni stazionarie (o punti fissi) della (258) sono le soluzioni della:

$$x = x^2 \tag{259}$$

(con $f(x) = x^2$) ovvero:

1. $x_1^* = 0$,
2. $x_2^* = 1$.

Ricavate le soluzioni stazionarie se ne esamina la stabilità. Per fare ciò si nota che la (258) equivale al seguente sistema (vedi la figura 96):

$$\begin{cases} x_{t+1} = x_t \\ x_{t+1} = x_t^2 \end{cases} \tag{260}$$

Se si considera il punto $x_2^* = 1$ e si linearizza in un intorno del punto la $f(x) = x^2$ (considerando che si ha $f'(x) = 2x$) utilizzando lo sviluppo di Taylor arrestato al primo ordine:

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \Delta((x - x_o)^2) \tag{261}$$

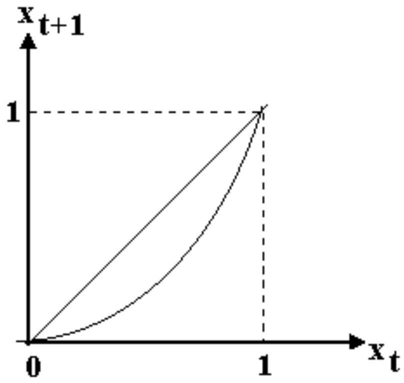

Figure 96: *Esempio di equazione non lineare (dettagli nel testo)*

si vede come sia il valore della derivata nel punto considerato a determinare il tipo di equilibrio (stabile o instabile). Nel nostro caso in $x_2^* = 1$ la derivata assume un valore in modulo maggiore di 1 per cui il punto di equilibrio considerato è di equilibrio **instabile** mentre in $x_1^* = 0$ la derivata ha valore nullo per cui tale punto è di equilibrio **stabile**. Ricapitolando si ha che:

1. $x_1^* = 0$ è un **attrattore**,
2. $x_2^* = 1$ è un **repulsore**.

Se si ricava la soluzione in forma chiusa della (258) si ha che vale:

$$x_t = x_0^{2t} = (x_0^2)^t \tag{262}$$

in modo che sia possibile individuare il bacino di attrazione imponendo:

$$x_0^2 < 1 \tag{263}$$

e quindi:

$$-1 < x_0 < 1 \tag{264}$$

ovvero se il punto iniziale si trova fra tali estremi allora la soluzione converge asintoticamente (ovvero per $t \longrightarrow +\infty$) alla soluzione di equilibrio stabile $x_1^* = 0$.
Vediamo un altro esempio in $\mathbb{R}^2$ (vedi la figura 97). Supponiamo di avere:

$$x_{t+1} = \sqrt{x_t} \tag{265}$$

Le soluzioni stazionarie le trovo risolvendo la seguente equazione:

$$x = \sqrt{x} \tag{266}$$

(con $f(x) = \sqrt{x}$) in modo che siano:

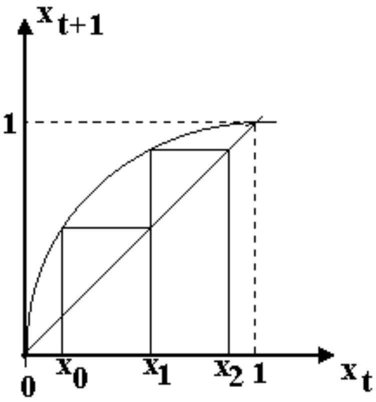

Figure 97: *Altro esempio di equazione non lineare (dettagli nel testo)*

1. $x_1^* = 0$,
2. $x_2^* = 1$.

Se calcolo la derivata della f ottengo:

$$f'(x) = \frac{1}{2\sqrt{x}} \tag{267}$$

che non è utilizzabile per valutare la stabilità della $x_1^* = 0$ (dal momento che non è definita in tale punto) mentre per la $x_2^* = 1$ ci dice che tale punto è **globalmente stabile** in quanto (vedi la figura 97) è facile verificare graficamente come, per ogni punto iniziale $x_0 > 0$, l'evoluzione del sistema faccia convergere le traiettorie verso il punto 1 sull'asse delle ascisse.

A.1.12 Modelli economici

Vediamo, in questa sezione, alcuni esempi di modelli economici descritti da coppie di equazioni alle differenze del primo ordine del tipo:

$$\begin{cases} x_{t+1} = ax_t + by_t + e \\ y_{t+1} = cx_t + dy_t + f \end{cases} \tag{268}$$

ovvero:

$$\begin{pmatrix} x_{t+1} \\ y_{t+1} \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_t \\ y_t \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix} \tag{269}$$

che in forma matriciale diventa:

$$X_{t+1} = AX_t + B \tag{270}$$

Gli autovalori della (270) li posso trovare risolvendo l'equazione caratteristica associata che ha la forma seguente:

$$\lambda^2 - tr(A)\lambda + det(A) = 0 \tag{271}$$

con:

1. $tr(A) = a + d$, traccia della matrice A;
2. $det(A) = ad - bc$, determinante della matrice A.

Il sistema (268) descrive anche un sistema non lineare dopo una operazione di linearizzazione nell'intorno di un punto. In questo caso, se f e g sono due funzioni non lineari, si ha:

$$\begin{cases} x_{t+1} = f(x_t, y_t) \\ y_{t+1} = g(x_t, y_t) \end{cases} \tag{272}$$

per cui posso usare lo Jacobiano:

$$J = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{pmatrix} \tag{273}$$

le cui righe sono i gradienti delle funzioni f e g. Se si risolve il sistema:

$$\begin{cases} x = f(x, y) \\ y = g(x, y) \end{cases} \tag{274}$$

si ricava la soluzione particolare del sistema non lineare (272) e si può valutare lo Jacobiano in tale punto in modo da ottenere una approssimazione lineare di tale sistema. Di nuovo si studiano gli autovalori come soluzioni della:

$$\lambda^2 - tr(A)\lambda + det(A) = 0 \tag{275}$$

Vediamo un esempio.
Si abbia il seguente sistema di equazioni alle differenze del primo ordine:

$$\begin{cases} x_{t+1} = x_t^2 + y_t \\ y_{t+1} = y_t^2 \end{cases} \tag{276}$$

Le soluzioni stazionarie del sistema (276) le calcolo imponendo $x_t = costante$ e $y_t = costante$ per ogni t ed ottenendo:

$$\begin{cases} x = x^2 + y \\ y = y^2 \end{cases} \tag{277}$$

Il sistema (277) ha le seguenti soluzioni:

1. $y = 0$ cui corrispondono $x = 0$ e $x = 1$,

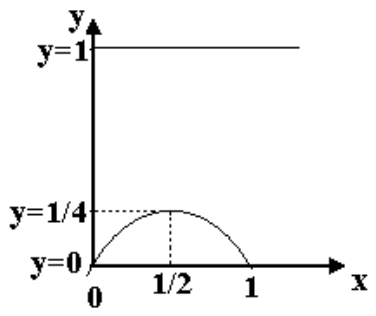

Figure 98: *Altro esempio di sistema di equazioni non lineari (dettagli nel testo)*

2. $y = 1$ cui non corrisponde nessuna soluzione reale per x (una soluzione in $\mathbb{R}$ per x dovrebbe essere una soluzione in $\mathbb{R}$ di $x = x^2 + 1$).

La figura 98 illustra la parabola di equazione $y = x - x^2$ e le rette $y = 0$ e $y = 1$ soluzioni di $y = y^2$.
Se riprendo il sistema:

$$\begin{cases} x = x^2 + y \\ y = y^2 \end{cases} \tag{278}$$

che è del tipo:

$$\begin{cases} x = f(x, y) \\ y = g(x, y) \end{cases} \tag{279}$$

e calcolo lo Jacobiano delle (278), per analizzare la stabilità, ottengo:

$$J = \begin{pmatrix} 2x & 1 \\ 0 & 2y \end{pmatrix} \tag{280}$$

in modo che, se valuto lo Jacobiano nei punti delle soluzioni stazionarie, ho i casi seguenti.

1. $$J(0, 0) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \tag{281}$$

   in modo da avere $tr(J) = 0$ e $det(J) = 0$. In questo caso (dalla (275) ho $\lambda_1 = \lambda_2 = 0$ ossia gli autovalori sono coincidenti per cui non è certo che la matrice sia diagonalizzabile sebbene il punto $(0, 0)$ sia di equilibrio stabile.

2. $$J(1, 0) = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \tag{282}$$

   in modo che sia $tr(J) = 1$ e $det(J) = 0$. In questo caso l'equazione caratteristica risulta essere:

   $$\lambda^2 - \lambda = 0 \tag{283}$$

con soluzioni:

(a) $\lambda_1 = 0$,

(b) $\lambda_2 = 1$

per cui:

(a) il punto $(1, 0)$ è localmente stabile dal momento che gli autovalori sono in modulo uno minore e uno uguale a 1;

(b) la matrice dello Jacobiano è diagonalizzabile dal momento che gli autovalori sono distinti.

A.1.13 Sistemi di equazioni differenziali lineari

Si suppone di avere il seguente sistema autonomo (in cui non compare esplicitamente t):

$$X'(t) = AX(t) \tag{284}$$

con:

1. $X(t) \in \mathbb{R}^n$,
2. $A \in \mathbb{R}^{n \times n}$ diagonalizzabile ovvero invertibile.

In questo contesto mi aspetto:

1. n autovalori distinti $\lambda_1, \ldots, \lambda_n$;
2. n autovettori distinti e linearmente indipendenti $v_1, \ldots, v_n$.

Il sistema (284) è di tipo omogeneo (dal momento che $b = 0$, $b \in \mathbb{R}^n$) ed ammette come soluzione la seguente:

$$X(t) = \sum_{i=1}^{n} c_i e^{\lambda_i t} v_i \tag{285}$$

Se invece si ha:

$$X'(t) = AX(t) + b \tag{286}$$

la soluzione generale è data dalla somma della soluzione dell'omogenea associata (ovvero la (285)) e della soluzione particolare ottenuta dalla (286) imponendo $X(t) = \textit{costante}$ in modo che sia:

1. $X' = 0$;
2. $AX = -b$

e quindi:

$$X^* = -A^{-1}b \tag{287}$$

La soluzione generale del sistema (286) risulta pertanto essere:

$$X(t) = \sum_{i=1}^{n} c_i e^{\lambda_i t} v_i - A^{-1}b \tag{288}$$

in cui i v_i sono autovettori linearmente indipendenti e compare la soluzione stazionaria $X^* = -A^{-1}b$.

A.1.14 Analisi di stabilità

Diamo ora alcuni cenni in merito alla stabilità della soluzione generale trovata:

$$X(t) = \sum_{i=1}^{n} c_i e^{\lambda_i t} v_i - A^{-1}b \tag{289}$$

Si hanno i casi seguenti:

1. se per ogni i $\lambda_i > 0$ il punto X^* è un **repulsore** ovvero è di **equilibrio instabile**;

2. se per ogni i $\lambda_i < 0$ il punto X^* è un **attrattore** ovvero è di **equilibrio stabile**;

3. se per $i = 1, \ldots, k$ $\lambda_i > 0$ e per $i = k+1, \ldots, n$ $\lambda_i < 0$ X^* è un punto di sella ovvero si hanno orbite di equilibrio (o varietà stabili) e altre orbite lungo le quali mi allontano dall'equilibrio come accade nel caso discreto (vedi, al proposito, la figura 93). Nel caso di $\mathbb{R}^2$ le varietà stabili sono elementi di $\mathbb{R}$ (ovvero rette) mentre in $\mathbb{R}^3$ le varietà stabili sono elementi di $\mathbb{R}^2$ ovvero piani. Identiche considerazioni possono essere fatte per le varietà instabili

Se siamo in $\mathbb{R}^2$ si hanno due autovalori λ_1 e λ_2 i cui segni possono essere:

1. concordi e positivi;

2. discordi;

3. concordi e negativi.

A.1.15 Caso di un sistema non lineare

Se la matrice A è diagonalizzabile (ovvero se gli autovalori sono tutti distinti o le loro molteplicità algebriche coincidono con le molteplicità geometriche) e se per ogni i si ha $\lambda_i \neq 0$ si ha che a livello locale in un intorno di X^* le dinamiche del sistema non lineare sono topologicamente equivalenti a quelle del sistema linearizzato associato. Si parla, in questo caso, di **attrattori di Lorentz**. Vediamo un esempio, nel caso continuo, di sistema non lineare di due equazioni:

$$\begin{cases} \dot{x} = f(x,y) = x^2 - y \\ \dot{y} = g(x,y) = \sqrt{x} - y \end{cases} \tag{290}$$

La soluzione di equilibrio del sistema (290) la calcolo imponendo:

$$\begin{cases} \dot{x} = 0 \\ \dot{y} = 0 \end{cases} \tag{291}$$

in modo da avere:

$$\begin{cases} x^2 - y = 0 \\ \sqrt{x} - y = 0 \end{cases} \tag{292}$$

ovvero:

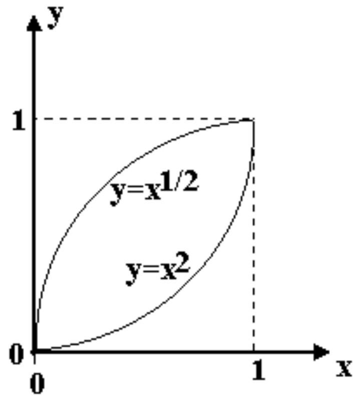

Figure 99: *Ulteriore esempio di sistema di equazioni non lineari (dettagli nel testo)*

$$\begin{cases} y = x^2 \\ y = \sqrt{x} \end{cases} \tag{293}$$

Le soluzioni del sistema (293) sono (vedi la figura 99):

1. $(0,0)$ in cui fallisce la linearizzazione dal momento che in tale punto la derivata della $\sqrt{x}$ non è definita;

2. $(1,1)$.

Se calcolo lo Jacobiano delle (290) ottengo (considerando la struttura dello Jacobiano mostrata dalla (273):

$$J = \begin{pmatrix} 2x & -1 \\ \frac{1}{2\sqrt{x}} & -1 \end{pmatrix} \tag{294}$$

e nel punto $(1,1)$ ho:

$$J = \begin{pmatrix} 2 & -1 \\ \frac{1}{2} & -1 \end{pmatrix} \tag{295}$$

con:

1. $det(J) = -\frac{3}{2}$;
2. $tr(J) = 1$,

In base a tali valori determino l'equazione caratteristica:

$$\lambda^2 - \lambda - \frac{3}{2} = 0 \tag{296}$$

Essendo $\Delta > 0$ e $\lambda_1\lambda_2 = -\frac{3}{2} < 0$ i due autovalori sono reali, distinti e di segno discorde per cui il punto $(1,1)$ è un punto di **sella**. Nel caso si trovassero autovalori complessi coniugati si deve fare l'analisi della parte reale dei singoli autovalori. Se tutti gli autovalori hanno parte reale negativa il punto di equilibrio X^* è un attrattore ovvero è di equilibrio **stabile** mentre se tutti gli autovalori hanno parte reale positiva il punto di equilibrio X^* è un repulsore ovvero è di equilibrio **instabile**. La figura 100 illustra i diversi casi che si possono verificare

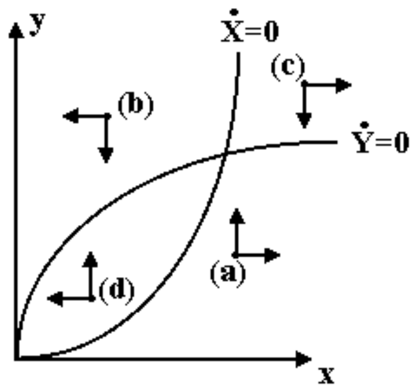

Figure 100: *Direzioni di crescita e decrescita (dettagli nel testo)*

relativamente al sistema (290). In figura sono illustrati i luoghi dei punti con $\dot{x} = 0$ (cui corrisponde $x^2 - y = 0$) e $\dot{y} = 0$ (cui corrisponde $\sqrt{x} - y = 0$). Si hanno i casi seguenti (vedi la figura 100).

1. **caso (a)**: in questo caso si ha:
   (a) $y < x^2$ per cui $\dot{x} > 0$;
   (b) $y < \sqrt{x}$ per cui $\dot{y} > 0$;

   e quindi entrambe le variabili crescono.

2. **caso (b)**: in questo caso si ha:
   (a) $y > x^2$ per cui $\dot{x} < 0$;
   (b) $y > \sqrt{x}$ per cui $\dot{y} < 0$;

   e quindi entrambe le variabili decrescono.

3. **caso (c)**: in questo caso si ha:
   (a) $y < x^2$ per cui $\dot{x} > 0$;
   (b) $y > \sqrt{x}$ per cui $\dot{y} < 0$;

   e quindi una delle variabile cresce e l'altra decresce.

4. **caso (d)**: in questo caso si ha:
   (a) $y > x^2$ per cui $\dot{x} < 0$;
   (b) $y < \sqrt{x}$ per cui $\dot{y} > 0$;

   e quindi una delle variabile cresce e l'altra decresce.

A.1.16 Ancora sulle equazioni alle differenze

Vediamo come si può eseguire una analisi qualitativa di una equazione alle differenze finite per appurare ([Cam04]):

1. se esistono punti di equilibrio,
2. se tali punti sono stabili o meno.

Nel seguito prenderemo in considerazione una equazione alle differenze finite del tipo:

$$x_{t+1} = f(x_t) \tag{297}$$

con $t = 0, 1, \ldots$. La f in genere non lineare in x_t si dice **funzione generatrice** perchè genera la successione dei valori a partire dal valore iniziale x_0. Una soluzione in forma chiusa della (297) è la seguente:

$$x_t = f^t(x_0) \tag{298}$$

dove f^t si dice iterata $t-$esima della f.
Un punto di equilibrio è una successione costante che verifica la (297) per cui $x_t = x^*$ è un punto di equilibrio della (297) sse si ha:

$$f(x^*) = x^* \tag{299}$$

ovvero x^* è un punto fisso della f.

Esempio A.10 *Si abbia:*

$$x_{t+1} = 4x_t(1 - x_t) \tag{300}$$

Si vuole x^* *tale che* $f(x^*) = x^*$ *ovvero si deve risolvere:*

$$x^* = 4x^*(1 - x^*) \tag{301}$$

ottenendo:

1. $x_1^* = 0$
2. $x_2^* = 3/4$

che sono punti di equilibrio per ogni t *dal momento che per entrambe le soluzioni si ha* $x_{t+1} = x_t$.

Data la:

$$x_{t+1} = f(x_t) \tag{302}$$

si possono avere più punti di equilibrio x_i^*, come mostra anche l'esempio precedente. In merito alla stabilità si ha:

1. se il valore iniziale x_0 è vicino ad uno dei punti x_i^* in modo che la successione $f(x_t)$ converga a x_i^* si dice che x_i^* è un punto di equilibrio **localmente asintoticamente stabile** ovvero, in modo più formale, x_i^* è localmente asintoticamente stabile se esiste $\varepsilon > 0$ tale che $\mid x_o - x_i^* \mid < \varepsilon \Rightarrow f^t(x_0) \longrightarrow x_i^*$ per $t \longrightarrow +\infty$;
2. l'insieme degli x_0 per cui ciò vale si dice **bacino di attrazione** di x_i^*.

Esempio A.11 *Si abbia:*

$$x_{t+1} = x_t^2 \tag{303}$$

La $x = f(x) = x^2$ *ha due punti di equilibrio:*

1. $x_1^* = 0$,
2. $x_2^* = 1$

Si vede facilmente che:

1. *se* $x_0 \in (-1, 1)$ *allora* $x_t = x_0^{2t} \longrightarrow 0$ *per* $t \longrightarrow +\infty$;
2. *se* $x_0 \notin (-1, 1)$ *allora* $x_t = x_0^{2t} \longrightarrow +\infty$ *per* $t \longrightarrow +\infty$;

per cui:

1. $x_1^* = 0$ *è localmente asintoticamente stabile* ,
2. $x_2^* = 1$ *è instabile.*

Se la successione $f^t(x_0)$ non converge ad un punto di equilibrio x^* ma rimane confinata in un intorno di tale punto si dice che x^* è un punto di equilibrio localmente stabile.

Esempio A.12 *Se si ha:*

$$x_{t+1} = -x_t \tag{304}$$

il solo punto di equilibrio è $x^* = 0$. Se si ha $x_0 = a$ la successione è la seguente:

$$a, -a, a, -a \ldots \tag{305}$$

che rimane confinata in un intorno di $x^* = 0$ che, pertanto, è un punto di equilibrio localmente stabile.

È facile capire che se un punto di equilibrio è globalmente asintoticamente stabile allora è anche localmente asintoticamente stabile ma che non vale il viceversa.
Un criterio per valutare la stabilità di un punto di equilibrio è il seguente.
Per la:

$$x_{t+1} = f(x_t) \tag{306}$$

se x^* è un punto di equilibrio soluzione della $x^* = f(x^*)$ si hanno due casi:

1. se $\mid f'(x^*) \mid < 1$ allora x^* si dice punto di equilibrio localmente asintoticamente stabile;

2. se $\mid f'(x^*) \mid > 1$ allora x^* si dice punto di equilibrio instabile.

Esempio A.13 *Se si ha:*

$$x_{t+1} = x_t^2 \tag{307}$$

si hanno due punti di equilibrio:

1. *$x_1^* = 0$ per il quale si ha $f'(x_1^* = 0 < 1$ in modo che $x_1^* = 0$ sia localmente asintoticamente stabile,*

2. *$x_2^* = 1$ per il quale si ha $f'(x_2^* = 2 > 1$ in modo che $x_2^* = 1$ sia instabile,*

A.1.17 Equazioni alle differenze finite lineari del secondo ordine

Si considera una equazione alle differenze finite del tipo ([Cam04]):

$$x_{t+2} + a_1(t)x_{t+1} + a_0(t) = b(t) \tag{308}$$

con:

1. $t = 0, 1, 2, \ldots,$

2. $a_i(t)$ a valori reali per $i = 1, 2,$

3. $b(t)$ a valori reali,

4. $a_0(t) \neq 0.$

Se $b(t) = 0$ la (308) si dice omogenea. La (308) ha una ed una sola soluzione per ogni coppia di valori iniziali x_0 e x_1. Data la (308) si considera l'equazione omogenea associata:

$$x_{t+2} + a_1(t)x_{t+1} + a_0(t) = 0 \tag{309}$$

L'insieme delle soluzioni W della (309) (se $a_0 \neq 0$) è uno spazio vettoriale di ordine 2 per cui:

1. la successione nulla appartiene a W;
2. se $y_t \in W$ allora $ky_t \in W$ per ogni $k \in \mathbb{R}$;
3. se $x_t, y_t \in W$ allora $x_t + y_t \in W$;
4. esistono due elementi in W non proporzionali fra di loro (ovvero linearmente indipendenti);
5. se tali elementi sono $y_t, z_t \in W$ allora ogni altra soluzione è ottenibile come loro combinazione lineare ovvero ogni x_t è scrivile come:

   $$x_t = \alpha y_t + \beta z_t \tag{310}$$

   con $\alpha, \beta \in \mathbb{R}$.

Data la (308) si ha che una sua soluzione generale è data dalla somma:

1. di una soluzione dell'equazione omogenea associata (309);
2. di una soluzione particolare della (308).

Nel caso i coefficienti che compaiono nella (308) siano costanti si parla di **equazioni alle differenze finite del secondo ordine a coefficienti costanti** che assumono la forma seguente:

$$x_{t+2} + a_1x_{t+1} + a_0 = b \tag{311}$$

cui è associata l'equazione omogenea:

$$x_{t+2} + a_1x_{t+1} + a_0 = 0 \tag{312}$$

Anche in questo caso, data la (311) si ha che una sua soluzione generale è data dalla somma:

1. di una soluzione dell'equazione omogenea associata (312);
2. di una soluzione particolare della (311).

In merito alla (312) si fa notare che, sulla base delle considerazioni fatte in merito allo spazio delle soluzioni W, per trovarne la soluzione si prendono in esame soluzioni del tipo:

$$x_t = \alpha^t \tag{313}$$

Sostituendo la (313) nella (312) e semplificando si ottiene la cosiddetta **equazione caratteristica** che ha la forma:

$$\alpha^2 - a_1\alpha + a_0 = 0 \tag{314}$$

per cui α^t è soluzione della (312) sse α è soluzione della (314). Se si pone:

$$\Delta = a_1^2 - 4a_0 \tag{315}$$

si hanno i casi seguenti.

1. $\Delta > 0$
   In questo caso la (314) ha due radici reali e distinte α_1 e α_2 per cui le soluzioni non proporzionali dell'omogenea sono α_1^t e α_2^t in modo che la soluzione dell'omogenea sia:
   $$x_t = c_1\alpha_1^t + c_2\alpha_2^t \tag{316}$$
   con c_1 e c_2 costanti da determinare.

2. $\Delta = 0$
   In questo caso la (314) ha una radice reale doppia α per cui le soluzioni non proporzionali dell'omogenea sono α^t e $t\alpha^t$ in modo che la soluzione dell'omogenea sia:
   $$x_t = c_1\alpha^t + c_2t\alpha^t \tag{317}$$
   con c_1 e c_2 costanti da determinare.

3. $\Delta < 0$
   In questo caso la (314) ha due radici complesse coniugate α_1 e α_2 per cui le soluzioni non proporzionali dell'omogenea sono α_1^t e α_2^t in modo che la soluzione dell'omogenea sia:
   $$x_t = c_1\alpha_1^t + c_2\alpha_2^t \tag{318}$$
   con c_1 e c_2 costanti da determinare. In questo caso si vuole che la soluzione sia in $\mathbb{R}$ per cui si procede come qui di seguito specificato. Si esplicitano le soluzioni α_1 e α_2:
   (a) $\alpha_1 = \rho(cos\theta + isen\theta) = \rho e^{i\theta}$,
   (b) $\alpha_1 = \rho(cos\theta - isen\theta) = \rho e^{-i\theta}$,

   in modo da scrivere:
   (a) $\alpha_1^t = \rho e^{it\theta} = \rho^t(cost\theta + isent\theta)$,
   (b) $\alpha_2^t = \rho e^{-it\theta} = \rho^t(cost\theta - isent\theta)$.

   I coefficienti c_1 e c_2 prendono valori in $\mathbb{C}$ per cui invece della:
   $$x_t = \rho^t[c_1(cost\theta + isent\theta) + c_2(cost\theta - isent\theta)] \tag{319}$$
   si possono trovare due soluzioni reali non proporzionali la cui somma dia luogo ad una x_t reale scegliendo in modo opportuno tali coefficienti in $\mathbb{C}$. Si ha:

(a) se $c_1 = c_2 = 1/2$ si ottiene $y_t = \rho^t cost\theta$,

(b) se $c_1 = 172i$ e $c_2 = -1/2i$ si ottiene $z_t = \rho^t sent\theta$,

in modo da ricavare:

$$x_t = k_1 y_t + k_2 z_t = \rho^t (k_1 cost\theta + k_2 sent\theta) \tag{320}$$

con $k_1, k_e \in \mathbb{R}$.

A questo punto si deve trovare una soluzione particolare dell'equazione non omogenea:

$$x_{t+2} + a_1 x_{t+1} + a_0 = b \tag{321}$$

Si esaminano brevemente solo i tre casi seguenti:

1. $b(t) = b$;
2. $b(t)$ polinomio anche non completo di grado n;
3. $b(t) = a\beta^t$.

1. In questo caso si hanno due sottocasi:

   (a) $1 + a_1 + a_0 \neq 0$, ovvero 1 non è soluzione dell'equazione caratteristica. La soluzione particolare cercata è supposta essere del tipo $x_t = k$ per cui sostituendo:
   $$k = \frac{b}{1 + a_1 + a_0} \tag{322}$$

   (b) $1 + a_1 + a_0 = 0$, ovvero 1 è soluzione dell'equazione caratteristica. La soluzione particolare cercata è supposta essere del tipo $x_t = tk$ per cui sostituendo e semplificando:
   $$k = \frac{b}{2 + a_1} \tag{323}$$

2. Supponiamo che $b(t)$ sia un polinomio anche non completo di grado n. Si hanno tre sottocasi:

   (a) se $\alpha = 1$ non è radice dell'equazione caratteristica si considera come soluzione la $x_t = P(t)$ con $P(t)$ polinomio completo di grado n;

   (b) se $\alpha = 1$ è radice semplice dell'equazione caratteristica si considera come soluzione la $x_t = tP(t)$ con $P(t)$ polinomio completo di grado n;

   (c) se $\alpha = 1$ è radice doppia dell'equazione caratteristica si considera come soluzione la $x_t = t^2 P(t)$ con $P(t)$ polinomio completo di grado n.

A questo punto si devono determinare i coefficienti incogniti delle potenze del polinomio $P(t)$ per cui si sostituisce il polinomio nella (320) e si sfrutta il principio di identità dei polinomi in base al quale i coefficienti dei termini di grado uguale devono essere uguali in modo da ricavare i coefficienti cercati.

3. Anche in questo caso si hanno tre sottocasi:

   (a) se β non è radice dell'equazione caratteristica si considera come soluzione la $x_t = k\beta^t$;

   (b) se β è radice semplice dell'equazione caratteristica si considera come soluzione la $x_t = tk\beta^t$;

   (c) se β è radice doppia dell'equazione caratteristica si considera come soluzione la $x_t = t^2k\beta^t$; Per il resto si procede in modo simile al precedente per ricavare k.

Vediamo ora alcune considerazioni in merito all'equilibrio e alla stabilità di equazioni lineari alle differenze finite a coefficienti costanti del tipo:

$$x_{t+2} + a_1 x_{t+1} + a_0 = b \tag{324}$$

Ogni soluzione costante, se esiste, $x_t = x^*$ è detta essere un **punto di equilibrio** della (324). Per l'esistenza di un punto di equilibrio della (324) si hanno i casi seguenti:

1. se 1 non è radice dell'equazione caratteristica ovvero se $1 + a_1 + a_0 \neq 0$ e $b \neq 0$ l'unico punto di equilibrio è:

$$x^* = \frac{b}{1 + a_1 + a_0} \tag{325}$$

2. se 1 è radice dell'equazione caratteristica ovvero se $1 + a_1 + a_0 = 0$ e $b \neq 0$ non si ha nessun punto di equilibrio;

3. se 1 è radice dell'equazione caratteristica ovvero se $1 + a_1 + a_0 = 0$ e $b = 0$ si hanno infiniti punti di equilibrio.

Nel seguito supporremo di essere nelle ipotesi del primo caso ovvero di avere un solo punto di equilibrio della della (324). Si esegue ora un'analisi della stabilità del punto di equilibrio considerando l'andamento della soluzione generale della (324) per $t \longrightarrow +\infty$. Per fare ciò si ricorre all'equazione caratteristica dell'equazione omogenea associata ovvero alla:

$$\alpha^2 + a_1\alpha + a_0 = 0 \tag{326}$$

le cui radici α_1 e α_2 possono essere:

1. reali e distinte;

2. reali e coincidenti;

3. complesse coniugate.

1. α_1 e α_2 sono reali e distinte per cui la soluzione generale della (324) ha la forma:
$$x_t = c_1\alpha_1^t + c_2\alpha_2^t + x^* \qquad (327)$$
in cui c_1 e c_2 vanno determinati in base alle condizioni iniziali. Si ha convergenza della soluzione generale alla soluzione particolare x^* (che è anche il punto di equilibrio) per $t \longrightarrow +\infty$ solo se:

   (a) $\mid \alpha_1 \mid < 1$,

   (b) $\mid \alpha_2 \mid < 1$.

   In questo caso x^* si dice soluzione globalmente asintoticamente stabile. Se $\mid \alpha_1 \mid > 1$ e/o $\mid \alpha_1 \mid > 1$ si ha instabilità. Se invece $\alpha_1 = -1$ e $\mid \alpha_2 \mid < 1$ o viceversa si ha una instabilità con oscillazioni attorno al valore di equilibrio x^*.

2. α_1 e α_2 sono reali e coincidenti ovvero $\alpha_1 = \alpha_2 = \alpha \neq 1$ la soluzione generale della (324) ha la forma:
$$x_t = \alpha^t(c_1 + tc_2) + x^* \qquad (328)$$
in cui c_1 e c_2 vanno determinati in base alle condizioni iniziali. Si ha convergenza della soluzione generale alla soluzione particolare x^* (che è anche il punto di equilibrio) per $t \longrightarrow +\infty$ solo se $\mid \alpha \mid < 1$. In ogni altro caso si ha instabilità, compreso il caso $\alpha = 1$.

3. α_1 e α_2 sono complesse coniugate del tipo $\rho(cos\theta \pm isen\theta)$ la soluzione generale della (324) ha la forma:
$$x_t = \rho^t(c_1 cost\theta + c_2 sent\theta) + x^* \qquad (329)$$
in cui c_1 e c_2 vanno determinati in base alle condizioni iniziali. Si ha convergenza della soluzione generale alla soluzione particolare x^* (che è anche il punto di equilibrio) per $t \longrightarrow +\infty$ solo se $\rho < 1$ mentre:

   (a) se $\rho > 1$ si ha instabilità;

   (b) se $\rho = 1$ si hanno oscillazioni.

   Il procedimento visto richiede che data una equazione del tipo:
$$x_{t+2} + a_1 x_{t+1} + a_0 = b \qquad (330)$$
se ne calcoli la soluzione generale per poter dire qualcosa sulla sua stabilità. Si ha un criterio che consente di stabilire come si comporta la (330) per $t \longrightarrow +\infty$ esaminando semplicemente i suoi coefficienti. Si ha che il punto

di equilibrio della (330) esiste ed è globalmente asintoticamente stabile sse sono verificate le condizioni seguenti:

$$\begin{cases} 1 - a_0 > 0 \\ 1 - a_1 + a_2 > 0 \\ 1 + a_1 + a_0 > 0 \end{cases} \tag{331}$$

A.1.18 Equazioni differenziali del primo ordine

Le equazioni differenziali del primo ordine sono equazioni in cui compare una funzione $x(t)$ con la sua derivata prima $x'(t)$ di grado anche superiore al primo ([Cam04]). Risolvere o integrare una equazione differenziale significa trovare tutte le funzioni $x(t)$ che la soddisfano ovvero che sono curve integrali o semplicemente integrali dell'equazione differenziale. Il caso più semplice è il seguente:

$$x'(t) = f(t) \tag{332}$$

la cui soluzione è:

$$x(t) = \int f(t)dt + c \tag{333}$$

si hanno infinite soluzioni, una per ogni valore della costante di integrazione c. Se è nota una condizione iniziale $x(t_0) = x_0$ si può ricavare c e pertanto si ha una soluzione unica. Nel caso di una equazione differenziale del secondo ordine si ha una soluzione unica sse sono note due condizioni iniziali.
Vediamo ora le equazioni differenziali lineari del primo ordine che hanno la forma seguente:

$$x'(t) = a(t)x(t) + b(t) \tag{334}$$

in cui:

1. $t \in I \subset \mathbb{R}$,
2. a e b sono funzioni continue in I.

Una funzione $x(t)$ è soluzione della (334) se la verifica e se è definita in I e ivi derivabile. Se $A(t)$ individua una primitiva di $a(t)$ in modo che sia $a(t) = A'(t)$ si può scrivere:

$$A(t) = \int a(t)\, dt \tag{335}$$

in modo che l'integrale generale della (334) sia:

$$x(t) = e^{A(t)} \int e^{-A(t)} b(t)\, dt \tag{336}$$

e si ha una soluzione unica in presenza di una condizione iniziale $x(t_0) = x_0$. Se si considera l'equazione omogenea associata alla (334) ovvero la:

$$x'(t) = a(t)x(t) \tag{337}$$

se ne può facilmente calcolare l'integrale generale:

$$x(t) = ce^{A(t)} \tag{338}$$

con $c \in \mathbb{R}$ e $A(t) = \int a(t)\, dt$. La soluzione della (334) può essere ottenuta come somma:

1. della soluzione dell'equazione omogenea associata;
2. di una soluzione particolare della (334).

Il caso più semplice da trattare è quello in cui nella (334) a e b sono costanti in modo che si abbia:

$$x'(t) = ax(t) + b \tag{339}$$

cui è associata la seguente equazione omogenea:

$$x'(t) = ax(t) \tag{340}$$

Tale equazione omogenea ha come soluzione la:

$$x(t) = ce^{at} \tag{341}$$

Come soluzione particolare della (339) si hanno i casi seguenti:

1. se $a \neq 0$ si può considerare come soluzione particolare della (339) la $x(t) = k$ in modo che sia:

   $$k = -\frac{b}{a} \tag{342}$$

   La soluzione generale della (339) risulta essere:

   $$x(t) = ce^{at} - \frac{b}{a} \tag{343}$$

   con c determinato grazie alla condizione iniziale $x(t_0) = x_0$;
2. se $a = 0$ per integrazione diretta si ottiene:

   $$x(t) = bt + c \tag{344}$$

   con c determinato grazie alla condizione iniziale $x(t_0) = x_0$.

Vediamo ora il caso in cui $b = b(t)$ per cui si ha:

$$x'(t) = ax(t) + b(t) \tag{345}$$

Si esaminano i due casi seguenti:

1. $b(t)$ polinomio anche non completo di grado n;
2. $b(t) = Be^{\alpha t}$.

1. In questo caso come soluzione particolare della (345) si considera un polinomio $Q(t)$ completo di grado n di cui vanno determinati i coefficienti. Si hanno due sottocasi:

   (a) se $a \neq 0$ la soluzione particolare è del tipo $x(t) = Q(t)$;

   (b) se $a = 0$ la soluzione particolare è del tipo $x(t) = tQ(t)$.

   Dopo aver fissato la forma della soluzione particolare la si sostituisce nella (345) e si applica il principio di identità dei polinomi in modo da ricavare i coefficienti incogniti cercati.

2. Si hanno due sottocasi:

   (a) se $\alpha \neq a$ si cerca una soluzione particolare del tipo $ke^{\alpha t}$;

   (b) se $\alpha = a$ si cerca una soluzione particolare del tipo $kte^{\alpha t}$.

   Fatto ciò si sostituisce la soluzione nella (345) e si determina il coefficiente k incognito.

Vediamo a questo punto alcune considerazioni in merito :

1. all'esistenza dei punti di equilibrio;

2. alla loro stabilità.

Un punto di equilibrio della:

$$x'(t) = ax(t) + b \tag{346}$$

è una sua soluzione costante. Per trovarla si sostituisce $x(t) = k$ nella (346) in modo che, se $a \neq 0$ si abbia:

$$k = -\frac{b}{a} \tag{347}$$

La soluzione generale della (346) risulta essere:

$$x(t) = ce^{at} - \frac{b}{a} \tag{348}$$

come somma della soluzione dell'equazione omogenea associata e di una soluzione particolare della (346). Il punto di equilibrio definito dalla (346) è globalmente asintoticamente stabile sse:

$$lim_{t \longrightarrow +\infty} x(t) = -\frac{b}{a} \tag{349}$$

ovvero sse $a < 0$. Se $a = 0$ la (346) non ammette punto di equilibrio (se $b \neq 0$). Per integrazione diretta è facile vedere che la soluzione risulta essere:

$$x(t) = bt + c \tag{350}$$

che diverge per $t \longrightarrow +\infty$.

A.1.19 Modello della ragnatela in tempo continuo

Si considerano la domanda $D(t)$ e l'offerta $S(t)$ di un singolo bene sul mercato ([Cam04]). Si suppone che le equazioni che le governano siano le seguenti:

$$D(t) = \alpha + ap(t) + a_1 p'(t) \tag{351}$$

(con $\alpha > 0$ domanda del bene gratuito, $a < 0$ e $a_1 < 0$)

$$S(t) = \beta + bp(t) \tag{352}$$

con $\beta > 0$, offerta del bene gratuito, e $b > 0$. I prezzi sono fissati in modo tale che, istante per istante, si abbia uno svuotamento del mercato (in inglese "market clearing") in modo che sia:

$$D(t) = S(t) \tag{353}$$

Sostituendo le (351) e (352) nella (353) si ha:

$$\alpha + ap(t) + a_1 p'(t) = \beta + bp(t) \tag{354}$$

ovvero:

$$p'(t) = \frac{b-a}{a_1} p(t) + \frac{\beta - \alpha}{a_1} \tag{355}$$

Per i segni dei parametri si ha che:

$$\frac{b-a}{a_1} < 0 \tag{356}$$

Si vuole l'andamento di $p(t)$. Come prezzo di equilibrio, soluzione particolare della (355), si ha:

$$p^* = \frac{\beta - \alpha}{b - a} \tag{357}$$

mentre la soluzione dell'omogenea è:

$$p(t) = ce^{\frac{b-a}{a_1} t} \tag{358}$$

in modo che la soluzione generale della (355) sia:

$$p(t) = ce^{\frac{b-a}{a_1} t} + p^* \tag{359}$$

in cui $c \in \mathbb{R}$ è determinato dalla condizione iniziale $p(t_0) = p_0$. Dal momento che vale la (356) si ha:

$$lim_{t \longrightarrow +\infty} p(t) = p^* \tag{360}$$

per cui si ha che esiste ed è unico un prezzo di equilibrio che è globalmente asintoticamente stabile.

A.1.20 Equazioni autonome

Se una equazione differenziale del primo ordine non è risolubile per integrazione si possono eseguire, in generale, analisi di tipo qualitativo per valutare se esiste o meno un punto di equilibrio e di che tipo ne sia la stabilità ([Cam04]). Una tale analisi di tipo qualitativo è possibile nel caso di equazioni differenziali autonome ovvero del tipo:

$$x'(t) = f(x(t)) \tag{361}$$

scritte come:

$$x' = f(x) \tag{362}$$

Se si fa l'ipotesi che per ogni $x_0 = x(t_0)$ la (362) abbia una sola soluzione in $[t_0, +\infty)$ si ha:

1. una soluzione costante $x(t) = k$ della (362) è un punto di equilibrio sse è tale che $f(k) = 0$ ovvero sse è uno zero della $f(x)$;
2. un punto di equilibrio della (362) è **globalmente asintoticamente stabile** sse:
$$lim_{t \longrightarrow +\infty} x(t) = k \tag{363}$$
3. un punto di equilibrio della (362) è **localmente asintoticamente stabile** sse esiste un intorno del punto di equilibrio tale che se $x(t)$ appartiene a tale intorno si ha:
$$lim_{t \longrightarrow +\infty} x(t) = k \tag{364}$$

Per l'analisi si può usare un **diagramma di fase** che:

1. in ascissa rappresenta la funzione $X = x(t)$;
2. in ordinata rappresenta la derivata $Y = x'(t)$;

in modo che la $x'(t) = f(x(t))$ sia rappresentata come $Y = f(X)$. In questo modo i punti di equilibrio della $x'(t) = f(x(t))$ sono gli zeri della $f(X)$ ovvero i punti in cui il grafico della Y incontra l'asse delle X. Dato uno di tali punti $X = a$ si hanno tre casi tipici:

1. se a sinistra del punto si ha $Y > 0$ si ha che la x' è positiva per cui la X cresce, se a destra si ha $Y < 0$ si ha che la x' è negativa per cui la X decresce e quindi il punto è di equilibrio stabile;
2. se a sinistra del punto si ha $Y < 0$ si ha che la x' è negativa per cui la X decresce, se a destra si ha $Y > 0$ si ha che la x' è positiva per cui la X cresce e quindi il punto è instabile;
3. si possono avere casi misti in cui si ha stabilità da un lato e instabilità dall'altro.

Riassumendo si ha che se in un'intorno del punto di attraversamento dell'asse delle ascisse la pendenza è negativa si ha equilibrio stabile se, invece, la pendenza è positiva si ha equilibrio instabile.

Esempio A.14 *Si abbia:*

$$x'(t) = x(t)^2 - 4x(t) \tag{365}$$

Se si pone $X = x(t)$ *e* $Y = x'(t)$ *si ha:*

$$Y = X^2 - 4X \tag{366}$$

Punti di intersezione con l'asse delle ascisse sono $X = 0$ *e* $X = 4$*. Poichè* $Y' = 2X - 4$ *si ha:*

1. *in* $X = 0$ $Y' = -4 < 0$ *per cui il punto è localmente asintoticamente stabile;*

2. *in* $X = 4$ $Y' = 8 > 0$ *per cui il punto è instabile.*

A.1.21 Equazioni differenziali lineari del secondo ordine

La forma generale è la seguente ([Cam04]):

$$x''(t) + a_1(t)x'(t) + a_0(t)x(t) = b(t) \tag{367}$$

in cui:

1. $t \in I \subset \mathbb{R}$

2. $a_1(t), a_0(t), b(t)$ sono funzioni a valori reali o complessi.

Se $b(t) = 0$ l'equazione (367) è detta omogenea. La (367) ha una ed una sola soluzione per ogni coppia di condizioni iniziali $x(t_0) = x_0$ e $x'(t_0) = y_0$.
Data l'equazione omogenea associata alla (367):

$$x''(t) + a_1(t)x'(t) + a_0(t)x(t) = 0 \tag{368}$$

l'insieme W delle soluzioni della (367) è uno spazio vettoriale di dimensione 2 per cui si ha:

1. $x(t) = 0$ è un elemento di W;

2. se $y(t) \in W$ allora $ky(t) \in W$ per ogni $\in \mathbb{R}$;

3. se $y(t) \in W$ e $z(t) \in W$ allora $y(t) + z(t) \in W$

4. in W esistono due soluzioni non proporzionali (ovvero linearmente indipendenti fra loro) $y(t)$ e $z(t)$ tali che ogni altra soluzione sia una loro combinazione lineare.

Data una equazione del tipo (367) una sua soluzione generale $x(t)$ è data dalla somma di:

1. una soluzione $y(t)$ dell'equazione omogenea associata (nella forma della (368));

2. una soluzione particolare $z(t)$ della (367);

ovvero si ha:

$$x(t) = y(t) + z(t) \tag{369}$$

Una versione semplice della (367) è quella in cui i coefficienti sono costanti:

$$x''(t) + a_1 x'(t) + a_0 x(t) = b(t) \tag{370}$$

cui è associata una equazione omogenea:

$$x''(t) + a_1 x'(t) + a_0 x(t) = 0 \tag{371}$$

Data una equazione del tipo (370) una sua soluzione generale $x(t)$ è data dalla somma di:

1. una soluzione $y(t)$ dell'equazione omogenea associata (nella forma della (371));
2. una soluzione particolare $z(t)$ della (370);

Considero l'equazione omogenea (371) e ne cerco una soluzione. Dato che mi servono delle soluzioni non proporzionali fra loro provo con $x(t) = e^{\lambda t}$ in modo che sia:

1. $x'(t) = \lambda e^{\lambda t}$,
2. $x''(t) = \lambda^2 e^{\lambda t}$.

Sostituendo tali relazioni nella (371) e semplificando si ottiene l'equazione caratteristica nella forma:

$$\lambda^2 + a_1 \lambda + a_0 = 0 \tag{372}$$

Se ora si pone $\Delta = a_1^2 - 4a_0$ si hanno tre casi:

1. $\Delta > 0$,
2. $\Delta = 0$,
3. $\Delta < 0$.

1. In questo caso la (372) ha due radici reali distinte λ_1 e λ_2 per cui l'equazione omogenea (371) ha la seguente soluzione:

   $$y(t) = c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 t} \tag{373}$$

   con $c_1, c_2 \in \mathbb{R}$ determinati in base alle condizioni iniziali.

2. In questo caso la (372) ha una radice reale doppia λ per cui l'equazione omogenea (371) ha la seguente soluzione:

   $$y(t) = (c_1 + tc_2)e^{\lambda t} \tag{374}$$

   con $c_1, c_2 \in \mathbb{R}$ determinati in base alle condizioni iniziali.

3. In questo caso la (372) ha due radici reali distinte λ_1 e λ_2 per cui l'equazione omogenea (371) ha la seguente soluzione:

$$y(t) = c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 t} \tag{375}$$

con $c_1, c_2 \in \mathbb{C}$. A noi interessano soluzioni in $\mathbb{R}$ per cui se:

(a) $\lambda_1 = \alpha + i\beta$,

(b) $\lambda_2 = \alpha - i\beta$,

si ha:

(a) $e^{\lambda_1 t} = e^{\alpha} e^{ibt} = e^{\alpha}(cosbt + isenbt)$,

(b) $e^{\lambda_2 t} = e^{\alpha} e^{-ibt} = e^{\alpha}(cosbt - isenbt)$

Sfruttando tali relazioni si ha:

(a) $\frac{e^{\lambda_1 t} + e^{\lambda_2 t}}{2} = e^{\alpha} cosbt$,

(b) $\frac{e^{\lambda_1 t} - e^{\lambda_2 t}}{2i} = e^{\alpha} senbt$

per cui l'equazione omogenea (371) ha la seguente soluzione:

$$y(t) = e^{\alpha t}(k_1 cosbt + k_2 senbt) \tag{376}$$

con $k_1, k_2 \in \mathbb{R}$ determinati in base alle condizioni iniziali.

A questo punto serve una soluzione particolare della:

$$x''(t) + a_1 x'(t) + a_0 x(t) = b(t) \tag{377}$$

Si esaminano tre casi:

1. $b(t) = b$;
2. $b(t)$ polinomio di grado n;
3. $b(t) = ae^{\alpha t}$

1. Si hanno tre sottocasi in funzione dei valori dei coefficienti della (377):
   (a) se $a_0 \neq 0$ come soluzione particolare si considera $z(t) = k$ e sostituendo nella (377) si ricava $k = \frac{b}{a_0}$;
   (b) se $a_0 = 0$ e $a_1 \neq 0$ come soluzione particolare si considera $z(t) = kt$ e sostituendo nella (377) si ricava $k = \frac{b}{a_1}$;
   (c) se $a_0 = 0$ e $a_1 = 0$ come soluzione particolare si considera $z(t) = kt^2$ e sostituendo nella (377) si ricava $k = \frac{b}{2}$;
2. In questo caso la soluzione appartiene alla classe dei polinomi completi $P(t)$ di grado n. Si hanno tre sottocasi:

(a) se $a_0 \neq 0$ si pone $z(t) = P(t)$;

(b) se $a_1 \neq 0$ ma $a_0 = 0$ si pone $z(t) = tP(t)$;

(c) se $a_1 = 0$ e $a_0 = 0$ si pone $z(t) = t^2 P(t)$;

A questo punto si sostituisce la $z(t)$ nella (377) e si usa il principio di identità dei polinomi per ricavare i coefficienti incogniti di $P(t)$.

3. Si hanno tre sottocasi in funzione del fatto che α sia o meno radice dell'equazione caratteristica:

   (a) se α non è radice dell'equazione caratteristica si ha $z(t) = ke^{\alpha t}$;

   (b) se α è radice semplice dell'equazione caratteristica si ha $z(t) = kte^{\alpha t}$;

   (c) se α è radice doppia dell'equazione caratteristica si ha $z(t) = kt^2 e^{\alpha t}$.

   A questo punto si sostituisce $z(t)$ nella (377) in modo da ricavare il valore di k.

Dopo aver visto come si risolve una equazione differenziale lineare del primo ordine vediamo qualcosa in merito a:

1. punto di equilibrio,

2. stabilità,

in relazione ad una equazione del tipo:

$$x''(t) + a_1 x'(t) + a_0 x(t) = b \tag{378}$$

Ogni soluzione costante (se esiste) della (378) è un punto di equilibrio. La (378) ha un punto di equilibrio sse la $x(t) = k$ soddisfa la (378) ovvero è tale che:

$$a_0 k = b \tag{379}$$

ovvero:

$$x^* = k = \frac{b}{a_0} \tag{380}$$

Nel caso della (378) si ha, quindi, che:

1. non si hanno punti di equilibrio se $a_0 = 0$ ma $b \neq 0$;

2. si hanno infiniti punti di equilibrio se $a_0 = 0$ e $b = 0$;

3. si ha un solo punto di equilibrio se $a_0 \neq 0$ e $b \neq 0$ e tale punto è dato dalla (380).

La condizione $a_0 = 0$ equivale a dire che $\lambda = 0$ è soluzione dell'equazione caratteristica. Limitatamente all'ultimo caso, per quanto riguarda la **stabilità** del punto di equilibrio si devono considerare le varie forme che può assumere la soluzione generale della (378). Ricordando quanto visto nei paragrafi precedenti si hanno i tre casi seguenti(considerando che la (380) ci dà la soluzione particolare).

1.
$$x(t) = c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 t} + x^* \tag{381}$$

in cui x^* è punto di equilibrio **globalmente asintoticamente stabile** sse $\lambda_1 < 0$ e $\lambda_2 < 0$, in ogni altro caso si ha instabilità. Sotto tali ipotesi si ha:
$$lim_{t \longrightarrow +\infty} x(t) = x^* \tag{382}$$

2.
$$x(t) = (c_1 + c_2 t)e^{\lambda t} + x^* \tag{383}$$

in cui x^* è punto di equilibrio **globalmente asintoticamente stabile** sse $\lambda < 0$, in ogni altro caso si ha instabilità. Sotto tali ipotesi si ha:

$$lim_{t \longrightarrow +\infty} x(t) = x^* \tag{384}$$

3.
$$x(t) = (c_1 cosbt + c_2 senbt)e^{\alpha t} + x^* \tag{385}$$

in cui x^* è punto di equilibrio **globalmente asintoticamente stabile** sse $\alpha < 0$, in ogni altro caso si ha instabilità. Sotto tali ipotesi si ha:

$$lim_{t \longrightarrow +\infty} x(t) = x^* \tag{386}$$

A.2 Elementi di integrazione numerica

A.2.1 Introduzione

In questa sezione vengono discussi brevemente alcuni degli algoritmi utilizzati per la risoluzione delle equazioni differenziali e per il calcolo degli integrali che compaiono nei modelli di simulazione dei sistemi. Nonostante si definiscano fra le varie variabili relazioni di tipo continuo, mediante l'uso di operatori derivata ed integrale, i metodi di risoluzione usati nella pratica ricorrono ad una discretizzazione di tali relazioni. In altre parti del testo abbiamo già visto quali forme possono assumere tali relazioni. In questa sezione si vuole vedere come, sulla base dei valori noti in intervalli di tempo distanziati ad una cadenza fissa, sia possibile determinare gli andamenti nel tempo di tali variabili. In questo contesto è facile capire come abbia una particolare importanza la cadenza con la quale si suppone vengano valutate le equazioni descrittive del modello.
Tali equazioni appartengono a tre categorie:

1. relazioni funzionali,
2. relazioni differenziali,
3. relazioni integrali.

Le relazioni della prima categoria caratterizzano le variabili ausiliarie, quelle della seconda le variabili di tipo flusso e quelle della terza le variabili di tipo livello.
Se, ad esempio, si ha una variabile ausiliaria X definita in funzione di altre variabili (costanti, ausiliarie, livelli, flussi, variabili ombra) in base ad una relazione funzionale del tipo:

$$X = f(Y) \tag{387}$$

in cui Y contiene tutte le variabili da cui la X dipende, una discretizzazione di tale relazione ha la forma seguente:

$$X_t = f(Y_t) \tag{388}$$

in cui il pedice t individua gli istanti in cui viene valutata tale relazione. Per la variabile t valgono i vincoli seguenti:

$$T_0 \leq t \leq T_{max} \tag{389}$$

in cui T_0 e T_{max} sono, rispettivamente, gli istanti iniziale e finale della simulazione. La variabile t viene incrementata, ad ogni passo della simulazione, di un valore Δt, sottomultiplo intero di $T_{max} - T_0$ ovvero

$$\Delta t = \frac{T_{max} - T_0}{n} \tag{390}$$

Una variabile di tipo flusso può essere descritta, in genere, da una relazione del tipo:

$$\frac{dX}{dt} = g(Y) \tag{391}$$

La versione discretizzata di tale relazione è la seguente:

$$\frac{\Delta X}{\Delta t} = g(Y_t) \tag{392}$$

In questo modo si ricava la variazione di una variabile flusso in un intervallo di tempo Δt:

$$\Delta X = g(Y_t)\Delta t \tag{393}$$

Tale relazione esprime la variazione ΔX come il prodotto di un valore $g(Y_t)$ (che si suppone non variare durante l'intervallo di tempo Δt) per la durata di tempo che diremo unitaria Δt. Se invece si vuole ricavare la variazione della variabile flusso fra due istanti successivi, considerando che[29] $\Delta X = X_{t+1} - X_t$, si ha:

$$X_{t+1} = X_t + g(Y_t)\Delta t \tag{394}$$

In questo modo si può ricavare l'andamento della variabile flusso nel tempo, dall'istante iniziale T_0 a quello finale $T_{max} = T_0 + n\Delta t$ della simulazione.

[29]Si fa notare che a rigore si dovrebbe scrivere $X_{t+\Delta t}$ e non X_{t+1}: la seconda notazione sebbene imprecisa è preferibile per la sua semplicità a patto che si abbia presente che l'istante $t+1$ si ottiene dall'istante t facendo scattare l'orologio di un intervallo unitario e indivisibile Δt.

In modo del tutto simile, l'equazione di bilancio per una variabile X di tipo livello ha la forma seguente:

$$X(t) = \int_{T_0}^{t} \varphi_{in}(\tau) - \varphi_{out}(\tau)\, d\tau \tag{395}$$

in cui φ_{in} rappresenta il contributo dei flussi in ingresso e φ_{out} rappresenta il contributo dei flussi in uscita. Da tale relazione, derivando rispetto a t, si ottiene:

$$\frac{dX}{dt} = \varphi_{in}(t) - \varphi_{out}(t) \tag{396}$$

supponendo che i valori dei flussi in T_0 siano nulli. In ogni caso è necessario che sia noto il valore iniziale (ovvero all'istante T_0) del livello. A questo punto, applicando il procedimento utilizzato per le variabili di tipo flusso, si ottiene:

$$X_{t+1} = X_t + (\varphi_{in}(t) - \varphi_{out}(t))\Delta t \tag{397}$$

in cui t ha il significato di una variabile discreta che varia per multipli interi di Δt.
Dalla discussione precedente dovrebbe essere evidente che, a parte il caso delle variabili ausiliarie, tutto si riduce alla determinazione di procedure di calcolo per la risoluzione di equazioni alle differenze finite del tipo:

$$X_{t+1} = X_t + (I_t - O_t)\Delta t \tag{398}$$

in cui I_t e O_t rappresentano, rispettivamente, il contributo in accrescimento e in decremento alla variabile X nell'intervallo Δt fra t e $t+1$, contributi supposti costanti su Δt. Per la risoluzione di tali equazioni esamineremo due tipi di algoritmi:

1. l'algoritmo di Eulero;
2. gli algoritmi di Runge-Kutta.

Tali algoritmi sono utilizzati dal risolutore inserito in $Vensim$.

A.2.2 Sviluppo in serie di Taylor

Prima di esaminare in dettaglio gli algoritmi di Eulero e di Runge-Kutta si richiamano brevemente alcune nozioni di base relative all'uso di serie di potenze per l'approssimazione di funzioni nell'intorno di un punto. In quanto segue si cerca di rispondere al seguente quesito: data una funzione $f(x)$, di una variabile reale x a valori reali, sotto quali condizioni è possibile rappresentare tale funzione come la somma di una serie di potenze? Collegato a tale quesito si ha il seguente: che forma ha tale rappresentazione?
Si ritiene utile osservare che:

1. se una funzione $f(x)$ può essere rappresentata sotto forma di una somma di una serie di potenze la si dice espandibile in una serie di potenze;

2. se una funzione è espandibile in una serie di potenze la si può sostituire con un polinomio (detto polinomio approssimante) che contiene solo i primi termini di tale serie;

3. tanto maggiore è il grado del polinomio approssimante tanto migliore è l'approssimazione e tanto minore è l'errore che si commette nel sostituire la funzione con il suo polinomio approssimante.

Per semplificare al massimo la trattazione, si assume che la funzione data $f(x)$ sia infinitamente differenziabile in un intorno del punto x_0 e si suppone che sia rappresentabile come la somma di una serie di potenze convergente alla funzione $f(x)$ in un intervallo che contiene il punto x_0. Sotto tali ipotesi si può scrivere:

$$f(x) = \sum_{i=0}^{\infty} a_i (x - x_0)^i \tag{399}$$

dove gli a_i sono coefficienti costanti (ovvero non dipendono da x).
Sfruttando le proprietà delle serie di potenze e la differenziabilità infinita della $f(x)$ in un intorno del punto x_0 si calcolano facilmente i valori dei coefficienti a_i. Vediamo il calcolo dei primi due, gli altri sono calcolati in modo analogo.
Se nella (399) si pone $x = x_0$ si ottiene:

$$a_0 = f(x_0) \tag{400}$$

Se si deriva una volta la (399) e si pone $x = x_0$ si ottiene:

$$a_1 = f'(x_0) \tag{401}$$

Ripetendo il procedimento si ottiene, per il coefficiente generico:

$$a_n = \frac{f^{(n)}(x_0)}{n!} \tag{402}$$

I coefficienti a_i così calcolati sono detti coefficienti di Taylor della serie:

$$f(x) = \sum_{i=0}^{\infty} \frac{f^{(i)}(x_0)}{i!} (x - x_0)^i \tag{403}$$

detta, a sua volta, sviluppo in serie di Taylor della funzione $f(x)$ in un intorno del punto x_0. Se la sommatoria che compare nella (403) è troncata ad un valore finito (pari a N) si parla di **polinomio di Taylor**. In tal caso la (403) può essere scritta come:

$$f(x) = \sum_{i=0}^{N} \frac{f^{(i)}(x_0)}{i!} (x - x_0)^i + R_N(x) \tag{404}$$

con $R_N(x)$ che rappresenta il resto ovvero la somma dei termini della serie di ordine superiore a N.

Prima di proseguire vediamo sotto quali condizioni la serie di Taylor (403) converge alla funzione $f(x)$ in un intervallo che contiene il punto x_0. Se si definisce il polinomio di Taylor di grado n:

$$T_n(x) = \sum_{i=0}^{n} \frac{f^{(i)}(x_0)}{i!}(x - x_0)^i \tag{405}$$

si può definire il resto $n-$esimo (o resto dopo n termini) $R_n(x)$ come:

$$R_n(x) = f(x) - T_n(x) \tag{406}$$

Da tali definizioni segue che la convergenza della serie di Taylor alla funzione $f(x)$ in un punto x equivale alla condizione:

$$lim_{n\to\infty} T_n(x) = f(x) \tag{407}$$

o, in modo equivalente:

$$lim_{n\to\infty} R_n(x) = 0 \tag{408}$$

Prima di chiudere la sezione si riprende la:

$$f(x) = \sum_{i=0}^{n} \frac{f^{(i)}(x_0)}{i!}(x - x_0)^i + R_n(x) \tag{409}$$

ovvero l'approssimazione di grado n della $f(x)$.
Se si pone:

$$x = x_0 + h \tag{410}$$

con $h \in \mathbb{R}$ piccolo a piacere, si ha:

$$f(x_0 + h) = \sum_{i=0}^{n} \frac{f^{(i)}(x_0)}{i!} h^i + R_n(x) \tag{411}$$

Se si pone $n = 1$ si ottiene il seguente sviluppo in serie:

$$f(x_0 + h) = f(x_0) + f'(x_0)h + R_2(h) \tag{412}$$

dove $R_2(h) = O(h^2)$ ovvero si commette, nel fare l'approssimazione, un errore di ordine h^2. La (412) si dice essere lo sviluppo in serie arrestato al primo ordine della $f(x)$: in base a tale relazione, in un intorno di x_0, si approssima la $f(x)$ con una retta passante per x_0 e avente pendenza $f'(x_0)$.
Se si pone $n = 2$ si ottiene il seguente sviluppo in serie:

$$f(x_0 + h) = f(x_0) + f'(x_0)h + \frac{f''(x_0)}{2!}h^2 + R_3(h) \tag{413}$$

dove $R_3(h) = O(h^3)$ ovvero si commette, nel fare l'approssimazione, un errore di ordine h^3. La (413) si dice essere lo sviluppo in serie arrestato al secondo ordine della $f(x)$: in base a tale relazione, in un intorno di x_0, si approssima la $f(x)$ con una funzione di secondo grado in h ovvero con una parabola.
Faremo uso di tali relazioni sia nel caso dell'algoritmo di Eulero sia nel caso degli algoritmi di Runge-Kutta.

A.2.3 Algoritmo di Eulero

Il metodo di Eulero ([Gal05], [GM80]) appartiene alla classe dei cosiddetti **metodi a un passo** per la risoluzione di equazioni differenziali ordinarie.
Il punto di partenza è una relazione del tipo:

$$\frac{dX(t)}{dt} = AX(t) \tag{414}$$

in cui $A = A(t)$ è un coefficiente in genere funzione del tempo. La (414) può essere facilmente risolta in forma chiusa[30]:

$$X(t) = X(0)e^{\int A dt} \tag{416}$$

In genere si pone:

$$A = N - M \tag{417}$$

in cui sia $N = N(t)$ sia $M = M(t)$ possono, in generale, essere funzioni del tempo e i cui valori sono noti per ogni t. Con tale posizione la (414) può essere scritta come:

$$\frac{dX(t)}{dt} = (N - M)X(t) \tag{418}$$

la cui soluzione in forma chiusa (nell'ipotesi semplificativa che sia N sia M siano indipendenti dal tempo) è esprimibile come:

$$X(t) = X(0)e^{(N-M)t} \tag{419}$$

La discretizzazione della equazione differenziale (414), come si ottiene facilmente, ha la forma seguente:

$$X(t + \Delta t) = X(t) + (N - M)X(t)\Delta t \tag{420}$$

che si può scrivere, usando la (418), come:

$$X(t + \Delta t) = X(t) + X'(t)\Delta t \tag{421}$$

se con $X'(t)$ si indica la derivata della $X(t)$ rispetto al tempo calcolata nel punto t. La (421) non è nient'altro che lo sviluppo in serie di Taylor arrestato al primo ordine della $X(t)$ (in cui il Δt gioca il ruolo di h nella (412)).
La situazione è illustrata in figura 101 in cui compaiono sia il valore "vero" di $X(t+\Delta t)$, posizionato sulla curva, sia il valore approssimato sulla retta tangente alla curva in $X(t)$, valore ottenuto con la discretizzazione. È intuitivo che tanto maggiore è la curvatura della curva tanto peggiore è l'approssimazione ovvero, più correttamente, tanto più piccolo è l'intorno di t in cui l'approssimazione è accettabile.

[30]Si scrive la (414) come:

$$\frac{dX(t)}{X(t)} = Adt \tag{415}$$

da cui $ln\, X(t) = \int Adt + K$ ovvero $X(t) = X(0)e^{\int Adt}$ con $X(0) = e^K$.

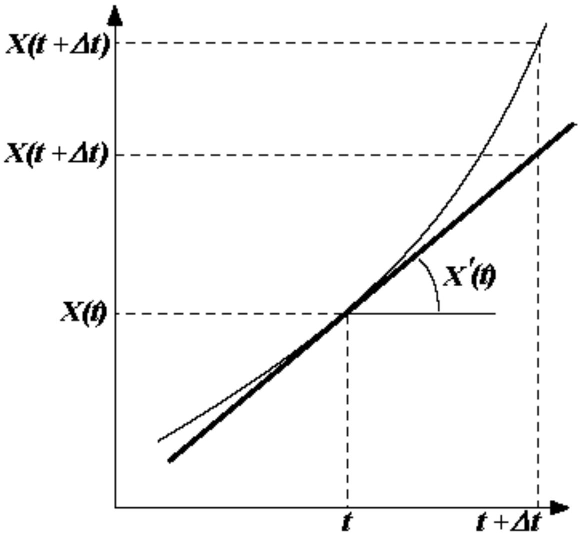

Figure 101: *Metodo di Eulero (da [Gal05]) (dettagli nel testo)*

È ovvio come nella (421) non sia possibile usare il valore esatto della $X'(t)$, che non è noto non essendo nota l'espressione della $X(t)$.
Il calcolo del valore della X in $t + \Delta t$ lo si fa, in pratica, usando la (420) sulla base del valore della X in t e dei valori, sempre in t, di N ed M.
Un primo modo per migliorare l'accuratezza del calcolo è quello di ridurre il valore dell'incremento Δt, ad esempio, dimezzandolo. Se si passa da Δt a $\Delta' t = \Delta t/2$, partendo da t, si ha che il valore successivo è calcolato in $t + \Delta' t$ sulla base del valore in t a cui si somma il contributo dato dal termine:

$$(N - M)X(t)\Delta' t \qquad (422)$$

In genere, tanto minore è il valore di Δt tanto migliore è l'accuratezza con cui si calcola la $X(t + \Delta t)$.
Si ricordi che alla:

$$X(t + \Delta t) = X(t) + (N - M)X(t)\Delta t = X(t) + NX(t)\Delta t - MX(t)\Delta t \qquad (423)$$

deve essere aggiunta la condizione iniziale $X(0)$ per $t = 0$.
La (423) ammette la seguente interpretazione, in cui non si tira in ballo il concetto di derivata. Il valore della X nel punto $t + \Delta t$ dipende:

1. dal valore della X in t;
2. dal valore del contributo istantaneo "in ingresso" $NX(t)$, considerato costante su Δt, pari a $NX(t)\Delta t$;
3. dal valore del contributo istantaneo "in uscita" $MX(t)$, considerato costante su Δt, pari a $MX(t)\Delta t$.

Da tale interpretazione si vede come, al tendere di Δt a 0, tranne che in casi patologici, l'accuratezza del calcolo non possa che migliorare.
Un modo per migliorare l'accuratezza del metodo di Eulero è quello di usare il valore della "derivata" in:

$$t + \frac{\Delta t}{2} \tag{424}$$

ovvero usare la:

$$X(t + \frac{\Delta t}{2}) = X(t) + (N - M)X(t)\frac{\Delta t}{2} \tag{425}$$

nella (423) in modo da ricavare:

$$X(t + \Delta t) = X(t) + (N - M)(X(t) + (N - M)X(t)\frac{\Delta t}{2})\Delta t \tag{426}$$

Per maggiori dettagli si rimanda alla sezione che segue.

A.2.4 Algoritmi di Runge Kutta

Si parla di algoritmi di Runge Kutta dal momento che ne esistono due, uno del secondo ordine e uno del quarto ordine, il solo usato in Vensim PLE.
L'algoritmo di Runge Kutta del secondo ordine ($RK2$) permette di ottenere approssimazioni migliori in $t + \Delta t$ utilizzando il valore della "derivata" calcolata in $t + \frac{\Delta t}{2}$ con il metodo di Eulero.
Da un punto di vista puramente analitico, il metodo sfrutta la[31]:

$$X(t + \Delta t) = X(t) + (N - M)X(t)\Delta t \tag{427}$$

in modo da calcolare il valore in $t + \Delta t$ in base al valore in $t + \frac{\Delta t}{2}$.
Per prima cosa, quindi, si calcola il valore della $X(t)$ in $\Delta t/2$ ovvero:

$$X(t + \frac{\Delta t}{2}) = X(t) + (N - M)X(t)\frac{\Delta t}{2} \tag{428}$$

poi si usa tale valore nella (427) in modo da avere:

$$X(t + \Delta t) = X(t) + (N - M)X(t + \frac{\Delta t}{2})\Delta t \tag{429}$$

che può essere riscritta come:

$$X(t + \Delta t) = X(t) + (N - M)(X(t) + (N - M)X(t)\frac{\Delta t}{2})\Delta t \tag{430}$$

ovvero:

$$X(t + \Delta t) = X(t) + (N - M)X(t)\Delta t + (N - M)^2 X(t)\frac{\Delta^2 t}{2} \tag{431}$$

[31] È ovvio che si tratta di una relazione approssimata per cui, a rigore, si dovrebbe usare un segno di $\simeq$ e non il segno di $=$.

in cui:

$$(N - M)X(t) \tag{432}$$

e:

$$(N - M)^2 X(t) \tag{433}$$

giocano, rispettivamente, il ruolo della $X'(t)$ e della $X''(t)$ nello sviluppo di Taylor arrestato al secondo ordine (vedi la (413)).
In questo modo si usa il valore della X in $t + \frac{\Delta t}{2}$ per calcolare il valore della X in $t + \Delta t$.
L'algoritmo di Runge Kutta del quarto ordine ($RK4$) applicato ad un intervallo $[t, t + \Delta t]$ sul quale si calcolano [Gal05] quattro "derivate" calcolate, rispettivamente:

1. una nel punto iniziale t;
2. due nel punto di mezzo $t + \Delta t/2$;
3. una nel punto terminale $t + \Delta t$.

In questo modo si ottengono quattro valori approssimati del valore cercato ovvero di $X(t + \Delta t)$. Tali valori vengono combinati fra di loro tramite una combinazione convessa (ovvero una combinazione lineare con coefficienti positivi la cui somma evale 1).
Il primo valore lo si ottiene applicando il metodo di Eulero su $[t, t + \Delta t]$ ovvero:

$$X_1(t + \Delta t) \simeq X(t) + (N - M)X(t)\Delta t \tag{434}$$

Il secondo valore lo si ottiene applicando il metodo di Eulero su $[t, t + \Delta t/2]$ ovvero:

$$X_1(t + \frac{\Delta t}{2}) \simeq X(t) + (N - M)X(t)\frac{\Delta t}{2} \tag{435}$$

che ci fornisce una seconda approssimazione di $X(t + \Delta t)$:

$$X_2(t + \Delta t) \simeq X(t) + (N - M)X_1(t + \frac{\Delta t}{2})\Delta t \tag{436}$$

La (436) ci consente di ottenere una nuova approssimazione di X in $t + \Delta/2$:

$$X_2(t + \frac{\Delta t}{2}) \simeq X(t) + (N - M)X_1(t + \frac{\Delta t}{2})\frac{\Delta t}{2} \tag{437}$$

che ci permette di nuovo di avere una ulteriore approssimazione della X in $t+\Delta t$:

$$X_3(t + \Delta t) \simeq X(t) + (N - M)X_2(t + \frac{\Delta t}{2})\Delta t \tag{438}$$

Con un meccanismo oramai noto si ricavano, infine, una ulteriore approssimazione della X in $t + \Delta t/2$:

$$X_3(t + \frac{\Delta t}{2}) \simeq X(t) + (N - M)X_2(t + \frac{\Delta t}{2})\frac{\Delta t}{2} \tag{439}$$

e una approssimazione della X in $t + \Delta t$:

$$X_4(t + \Delta t) \simeq X(t) + (N - M)X_3(t + \frac{\Delta t}{2})\Delta t \tag{440}$$

Le quattro approssimazioni calcolate in $t + \Delta t$ sono combinate mediante una combinazione convessa in modo da ottenere:

$$X(t + \Delta t) \simeq \frac{X_1(t + \Delta t)}{6} + \frac{X_2(t + \Delta t)}{3} + \frac{X_3(t + \Delta t)}{3} + \frac{X_4(t + \Delta t)}{6} \tag{441}$$

A.2.5 Gradi di approssimazione degli algoritmi

I metodi visti sono metodi approssimati di soluzione di equazioni differenziali ordinarie del primo ordine. Il **metodo di Eulero** permette di scrivere relazioni del tipo:

$$X(t + \Delta t) = X(t) + X'(t)\Delta t + O(\Delta^2 t) \tag{442}$$

che rappresentano approssimazioni del primo ordine nelle quali si commette un errore di $O(\Delta t^2)$ se si usa la relazione approssimata:

$$X(t + \Delta t) \simeq X(t) + X'(t)\Delta t \tag{443}$$

in base alla quale si approssima la curva con una retta.
L'**algoritmo di Runge Kutta** del secondo ordine ($RK2$) permette di ottenere una approssimazione del secondo ordine:

$$X(t + \Delta t) = X(t) + X'(t)\Delta t + \frac{X''(t)}{2}\Delta t^2 + O(\Delta^3 t) \tag{444}$$

che rappresenta una approssimazione del secondo ordine nella quale si commette un errore di $O(\Delta^3 t)$ se si usa la relazione approssimata:

$$X(t + \Delta t) \simeq X(t) + X'(t)\Delta t + \frac{X''(t)}{2}\Delta^2 t \tag{445}$$

in base alla quale si approssima la curva con una parabola.
È ovvio che non conoscendo i valori di $X'(t)$ e di $X''(t)$ si vuole procedere usando i valori noti di N ed M che contribuiscono alla approssimazione con i termini:

$$(N - M)X(t)\Delta t \tag{446}$$

e:

$$(N - M)X(t)\frac{\Delta^2 t}{2} \tag{447}$$

Vediamo come si possa procedere. Si parte dalla (445) scritta come:

$$X(t + \Delta t) \simeq X(t) + (X'(t) + \frac{X''(t)}{2}\Delta t)\Delta t \tag{448}$$

Se si applica Eulero per valutare X' in $t + \Delta/2$ si ottiene:

$$X'(t + \frac{\Delta t}{2}) \simeq X'(t) + X''(t)\frac{\Delta}{2} \tag{449}$$

e sostituendo nella (448) si ottiene (applicando di nuovo Eulero usando la derivata in $t + \Delta/2$):

$$X(t + \Delta t) \simeq X(t) + X'(t + \frac{\Delta t}{2})\Delta t = X(t) + (N - M)X(t + \frac{\Delta t}{2})\Delta t \quad (450)$$

sfruttando il fatto che $X'(t + \frac{\Delta t}{2}) = (N - M)X(t + \frac{\Delta t}{2})$. Se in più si impone nella (450):

$$X(t + \frac{\Delta t}{2}) \simeq X(t) + (N - M)X(t)\frac{\Delta t}{2} \quad (451)$$

si ottiene la versione effettiva della (445) ovvero:

$$X(t + \Delta t) \simeq X(t) + (N - M)(X(t) + (N - M)X(t)\frac{\Delta t}{2})\Delta t \quad (452)$$

che equivale ad uno sviluppo in serie di Taylor arrestato al secondo ordine.
Utilizzando un procedimento analogo si può vedere come il metodo di Runge-Kutta del quarto ordine equivalga ad una valutazione di $X(t + \Delta t)$ con uno sviluppo in serie di Taylor arrestato al quarto ordine in modo che sia:

$$X(t + \Delta t) = X(t) + X'(t)\Delta t + \frac{X''(t)}{2!}\Delta^2 t + \frac{X'''(t)}{3!}\Delta^3 t + \frac{X''''(t)}{4!}\Delta^4 t + O(\Delta^5 t) \quad (453)$$

A.2.6 I metodi di Eulero e Runge-Kutta e l'integrazione

In chiusura si riesaminano i metodi visti da un punto di vista più matematico e formale. Quanto segue deriva essenzialmente dal capitolo 16 di [PTVF92].
Il nostro obiettivo è la soluzione di equazioni differenziali ordinarie. Se tali equazioni sono di ordine superiore al primo abbiamo visto come le si possa trasformare in un sistema di equazioni differenziali del primo ordine. In tal modo la soluzione di una equazione differenziale ordinarie di ordine N è riconducibile alla soluzione di N equazioni differenziali ordinarie del primo ordine accoppiate nella forma:

$$\frac{dy_i(x)}{dx} = f_i(x, y_1, \ldots y_N) \quad (454)$$

con $i = 1, \ldots, N$ e dove le f_i sono note. Per completare la conoscenza del problema sono necessarie le **condizioni al contorno** definite in base al valore delle y_i in un qualche valore iniziale x_s. In questo caso lo scopo è valutare le y_i in un qualche punto finale x_f. Negli esempi visti la variabile indipendente è il tempo che varia fra un istante iniziale ed uno finale.
Il metodo di soluzione più semplice consiste nel riscrivere la (454) (scritta in una sola variabile) come:

$$y(x + \Delta x) = y(x) + f(x)\Delta x \quad (455)$$

Nei casi visti si ha $x = t$. La (455) descrive il **metodo di Eulero** e, nel caso si abbia un Δx sufficientemente piccolo, consente di ricavare i valori di y in

multipli interi di Δx con una buona approssimazione.
Il metodo di Eulero è importante perchè dalla sua idea di base discendono tutti gli altri: applicare piccoli incrementi ad una funzione che corrisponde ad una derivata moltiplicata per incrementi del passo Δx.
Metodi migliori in relazione al grado di approssimazione e alla stabilità in funzione del "passo" Δx sono i **metodi di Runge-Kutta** che:

1. propagano una soluzione su un intervallo,
2. combinano le informazioni derivanti da più applicazioni successive del metodo di Eulero,
3. usano le informazioni ottenute per approssimare uno sviluppo in serie di Taylor di un ordine superiore al primo.

I metodi di Runge-Kutta danno praticamente sempre risultati corretti ma non sono molto veloci se non nei casi in cui la valutazione della f non sia molto onerosa e si richiedano precisioni modeste.
Per prima cosa vediamo la formula di Eulero (per semplicità si usa h invece di Δx). Tale formula ha la forma seguente:

$$y_{n+1} = y_n + h\, f(x_n, y_n) \tag{456}$$

con $n = 0, \ldots, N$ e:

$$x_{n+1} \equiv x_n + h \tag{457}$$

con $n = 0, \ldots, N$. La (457) stabilisce la legge secondo la quale viene incrementata la variabile indipendente x. La (456) ha una struttura asimmetrica: la soluzione viene valutata su un intervallo di ampiezza h usando il valore della derivata all'inizio di tale intervallo. L'errore che si commette usando la (456) è esprimibile come $O(h^2)$ per cui il metodo non è molto accurato (soprattutto in confronto ad altri) e, in più, presenta problemi di stabilità in funzione del valore di h utilizzato.
Un uso possibile della (456) è quello di consentire la valutazione di un valore "tentativo" nel punto di mezzo dell'intervallo in modo da usare il valore di x e y in tale punto intermedio per valutare il passo effettivo su tutto l'intervallo. Formalmente si ha ([PTVF92]):

$$k_1 = hf(x_n, y_n) \tag{458}$$

$$k_2 = hf(x_n + \frac{1}{2}h, y_n + \frac{1}{2}k_1) \tag{459}$$

da cui:

$$y_{n+1} = y_n + k_2 + O(h^3) \tag{460}$$

in modo da ottenere un metodo del secondo ordine ovvero con un errore del tipo $O(h^3)$[32]. Le (458)÷(460) definiscono il metodo di Runge-Kutta del secondo ordine o del punto di mezzo.

[32]Un metodo è detto essere di ordine n se il suo errore è del tipo $O(h^{n+1})$.

Proseguendo su questa linea di ragionamento si può definire il metodo di Runge-Kutta del quarto ordine. Tale metodo sfrutta diversi coefficienti che sono combinati fra di loro in modo da definire una approssimazione di ordine superiore. Le relazioni che definiscono il metodo sono le seguenti:

$$\begin{cases} k_1 = hf(x_n, y_n) \\ \\ k_2 = hf(x_n + \frac{1}{2}h, y_n + \frac{1}{2}k_1) \\ \\ k_3 = hf(x_n + \frac{1}{2}h, y_n + \frac{1}{2}k_2) \\ \\ k_4 = hf(x_n + h, y_n + k_3) \\ \\ y_{n+1} = y_n + \frac{k_1}{6} + \frac{k_2}{3} + \frac{k_3}{3} + \frac{k_4}{6} + O(h^5) \end{cases} \tag{461}$$

Tale metodo richiede la valutazione delle (461). Si fa notare tuttavia come ad un metodo di ordine superiore non corrisponda sempre una maggiore accuratezza per cui se è vero che il metodo del quarto ordine è in genere superiore ad un metodo del secondo ordine tale affermazione non può essere estesa a tutti i casi.

References

[Cam04] Alberto Cambini. *Matematica per l'Economia.* Facoltà di Economia, materiali didattici, 2004.

[Gal05] Giorgio Gallo. *Appunti per il Corso di Modellistica Ambientale.* Dipartimento di Informatica, materiali didattici, 2005.

[GM80] Giuseppe Ghelardoni and Pietro Marzulli. *Argomenti di analisi numerica, vol. II.* ETS Università, 1980.

[Kir98] Craig W. Kirkwood. *System Dynamics Methods: A Quick Introduction.* College of Business, Arizona State University, 1998.

[PTVF92] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, 1992.

[RAD+83] Nancy Roberts, David Andersen, Ralph Deal, Michael Garet, and William Shaffer. *Introduction to Computer Simulation.* Addison Wesley Publishing Company, 1983.

[Ven05] Ventana Systems Inc. *Vensim PLE User's Guide*, 4 edition, 2005.