Funzione signum, notazioni di Iverson, e altro*

Luciano Battaia†

Sommario

Prendendo spunto dall'idea di discutere la definizione della funzione *signum*, in questo articolo presento alcune considerazioni sulla funzione caratteristica di un insieme, sulle notazioni di Iverson e su alcuni problemi collegati ai calcoli approssimati. Propongo inoltre alcuni considerazioni sulla risoluzione di disequazioni nei reali e sui problemi della continuità e derivabilità per le funzioni reali di variabile reale. Il taglio dell'articolo è prevalentemente didattico e ha solo un carattere introduttivo.

Indice

Introduzione

Le notazioni matematiche evolvono continuamente, come del resto succede a tutti i linguaggi. I matematici sono moderatamente conservatori, e questo impedisce cambi continui e troppo rapidi che poi si rivelerebbero inutili o controproducenti. D'altra parte

*Versione del 24 gennaio 2007.

†http://www.batmath.it

gli stessi matematici sono anche moderatamente innovatori e questo consente la continua semplificazione e "modernizzazione" delle simbologie usate.

Un esempio famoso, riguardante le notazioni, è costituito dall'uso delle parentesi come "contenitori": probabilmente sconosciute prima del *General trattato di numeri e misure* (1556-1560) di Niccolò Tartaglia, divennero di uso comune solo a seguito dell'opera di Eulero (1707-1783). Prima si usavano altri artifici, tra cui abbastanza comune era quello delle barre orizzontali che indicavano le espressioni da raggruppare. È chiaro che una simile notazione risulta assolutamente improponibile per espressioni complesse.

Il problema delle notazioni usate non è secondario in matematica: si pensi, per esempio, alle difficoltà di calcolo che presenta l'uso dei numeri scritti in un sistema non posizionale (come quello dei romani, per esempio).

Anche per quanto riguarda le definizioni adottate nei vari contesti ci possono, naturalmente, essere differenze, pur se quasi sempre riferite a situazioni marginali e di limitato interesse.

In questo articolo voglio occuparmi della funzione signum e delle diverse definizioni adottate. L'argomento è di per sè del tutto marginale, e in effetti è qui semplicemente preso a pretesto per trattare qualche altro problema, di gran lunga più interessante: il concetto di funzione caratteristica di un insieme, le notazioni di Iverson, il problema delle approssimazioni nei calcoli automatici, la risoluzione delle disequazioni, i concetti di continuità e derivabilità delle funzioni reali di variabile reale, ...

Sfogliando i testi più diffusi ed esaminando i software di calcolo (soprattutto simbolico) ho potuto trovare diverse definizioni della funzione signum.[1] Tutti concordano nel porre $\operatorname{sgn}(x) = 1$ se $x > 0$, $\operatorname{sgn}(x) = -1$ se $x < 0$. Le differenze si limitano alla situazione che si presenta nel punto 0, dove ho riscontrato le seguenti convenzioni possibili:

1. $\operatorname{sgn}(0) = 0$;
2. $\operatorname{sgn}(0)$ non è definita;
3. $\operatorname{sgn}(0) = 1$;
4. $\operatorname{sgn}(0) = -1$
5. $\operatorname{sgn}(0) = \pm 1$.

La scelta n.1 è fatta per esempio in Gelbaum e Olmsted (1979), Knuth (1992), Borowski e Borwein (2002) e in Lesina (1994). Quest'ultima citazione è importante perchè si tratta del più famoso manuale di stile in lingua italiana e le convenzioni riportate sono in stretto accordo con le normative ufficiali ISO (International Organization for Standardization). In Trombetta (2006) si considera esplicitamente una funzione f con i valori qui riportati, ma non ne viene dato un nome specifico. La scelta n.2 è fatta, tra l'altro, in Prodi (1970) e in Laforgia (1995). Meno diffuse le rimanenti. L'ultima delle convenzioni, formalmente scorretta dal punto di vista della definizione di funzione, è adottata in alcuni software di calcolo simbolico, per esempio in *Derive*,[2] e ha una sua giustificazione logica, come vedremo nel § 5.

Per quanto riguarda i software di calcolo simbolico più diffusi, tutti, con l'eccezione già considerata di *Derive*, fanno la scelta n.1; tra questi spiccano *Maxima*, *Mathematica*,

[1] La funzione signum è anche nota, nei paesi di lingua anglosassone, come *sign function*. Quest'ultimo nome è però da evitare, perchè facilmente confondibile, soprattutto nel parlato, con la funzione *seno* che, in inglese, è detta *sin function*. Il simbolo ufficialmente adottato dalla normativa ISO 31-11 è "sgn" (in carattere tondo come tutti nomi specifici di funzione). Mi atterrò ovviamente a questa notazione nel presente articolo.

[2] Tutti i nomi di software usati nel presente articolo sono copyright dei rispettivi produttori: *Derive*, *Maxima*, *Mathematica*, *Maple*, *Cabri*, *Excel*, *Access*.

Maple. Anche alcuni altri programmi di uso molto frequente, e che hanno bisogno di definire la funzione signum, seguono questa convenzione; tra gli altri *Cabri*, *Excel*, *Access*. Stessa situazione per i linguaggi di programmazione: Visual Basic, C++, Java, VBscript, Javascript, tutti adottano la scelta n.1.

1 Funzione caratteristica

Definizione 1.1. *Dato un sottoinsieme A di un insieme E, si chiama* ***funzione caratteristica*** *o anche* ***funzione indicatrice*** *di A in E la funzione $\chi_A\colon E \to \{0,1\}$ data da:*

$$\chi_A(x) \stackrel{\text{def}}{=} \begin{cases} 1 & \textit{se} \quad x \in A \\ 0 & \textit{se} \quad x \notin A \end{cases} .$$

È chiaro che ogni sottoinsieme A di E ha un ben precisa funzione caratteristica e, viceversa, che ad ogni funzione $f\colon E \to \{0,1\}$ corrisponde un ben determinato sottoinsieme A di E. Questo singifica che assegnare un insieme A equivale ad assegnarne la funzione caratteristica.

In sostanza questa funzione misura il *grado di appartenenza* di un elemento x di un insieme *universo* E ad un sottoinsieme A di E. Nel caso della teoria classica degli insiemi, di cui ci occupiamo qui, tale grado di appartenenza può essere solo 0 o 1: un elemento x ha solo due possibilità, o appartiene ad un sottoinsieme, o non vi appartiene.

Le cose vanno in maniera decisamente diversa in molte situazioni reali, nelle quali la teoria classica degli insiemi manifesta tutti i suoi limiti. Consideriamo un esempio per capirci meglio. Se in un concorso per piloti aereonautici è richiesta la statura minima di 1.80 m, un candidato che misuri 1.79 m non è (o almeno non dovrebbe!) essere ammesso: si tratta di una situazione gestibile dalla teoria classica degli insiemi. Se però ci mettiamo a considerare la statura delle persone da un altro punto di vista, e decidiamo di considerare *alti* tutti gli individui con statura di almeno 1.80 m, siamo naturalmente portati a *non* considerare *bassi* gli individui che abbiano una statura di 1.79 m: ci viene spontaneo pensare che abbiano un *grado di altezza* inferiore a quello richiesto per considerarli alti, ma che questo non sia sufficiente a considerarli bassi. Questa situazione non è gestibile dalla teoria classica degli insiemi: un individuo può essere alto in diversi gradi.

Da un punto di vista formale, questo porta all'introduzione di una nuova teoria degli insiemi: gli *insiemi fuzzy*. Anche questi insiemi possono essere definiti mediante una funzione caratteristica, la quale però non avrà solo due valori, come nella teoria classica. Precisamente la definizione che si dà è la seguente: un sottoinsieme A di un insieme universo E è definito dalla sua funzione caratteristica $\chi_A\colon E \to [0,1]$, funzione che associa ad ogni elemento x di E un qualunque numero reale compreso tra 0 e 1. Di un elemento x tale che $\chi_A(x) = 0$ si potrà dire che *non appartiene* ad A; di un elemento x tale che $0 < \chi_A(x) < 1$ si potrà dire che esso ha un *grado $\chi_A(x)$ di appartenenza* ad A; di un elemento x tale che $\chi_A(x) = 1$ si potrà dire semplicemente che *appartiene* ad A.

Tornando alla teoria classica degli insiemi vediamo due interessanti proprietà della funzione caratteristica di un insieme.

1. $\chi_{A\cap B}(x) = \min\big(\chi_A(x), \chi_B(x)\big) = \chi_A(x) \cdot \chi_B(x)$
2. $\chi_{A\cup B}(x) = \max\big(\chi_A(x), \chi_B(x)\big) = \chi_A(x) + \chi_B(x) - \chi_A(x) \cdot \chi_B(x)$

In particolare la seconda proprietà dice in sostanza che per fare l'unione di due insiemi si devono contare tutti gli elementi del primo e tutti quelli del secondo, e infine sottrarre tutti quelli in comune, altrimenti essi sarebbero contati due volte.

2 Notazione di Iverson

Kenneth Iverson, il padre di *APL*, introduce una notazione molto utile nella scrittura dei linguaggi di programmazione, ma vantaggiosa anche in altri contesti. Si tratta della notazione, nota come **Parentesi di Iverson**, che ora definiremo.

Definizione 2.1. *Sia P una proposizione che può avere solo due valori di verità:* vero *o* falso*. Allora*

$$[P] \overset{\text{def}}{=} \begin{cases} 1 & \textit{se} \quad P \textit{ è vera} \\ 0 & \textit{se} \quad P \textit{ è falsa} \end{cases}$$

Una interessante trattazione di questa notazione è contenuta in Knuth (1992), a cui rimando per approfondimenti.

Per esempio la funzione $f\colon \mathbb{R} \to \mathbb{R}$ definita da $f(x) = [x > 0]$ è la funzione che vale 1 sui reali positivi, 0 sui reali negativi e sullo zero.

Questa notazione consente di scrivere la famosa funzione δ di Kronecker in maniera molto elegante:

$$\delta_{ij} \overset{\text{def}}{=} [i = j], \quad \text{per } i,\, j \text{ interi arbitrari}$$

Un'altra applicazione molto utile di questa notazione è la possibilità di scrivere ogni somma come una somma infinita. Per esempio, invece di scrivere

$$\sum_{k=2}^{n} a_k\,,$$

si potrà scrivere

$$\sum_{k} a_k [2 \leq k \leq n]\,.$$

Anche se la modifica può sembrare più formale che sostanziale, in realtà il passaggio da una scrittura "su due linee" a una "su una sola linea" risulta molto vantaggiosa per la maggiore facilità di implementazione in un algoritmo di calcolo automatico.

Il fatto che molti software di calcolo simbolico implementino questa funzione, seppure con nomi diversi, è una dimostrazione della sua crescente importanza. *Mathematica*, per esempio, la definisce con il nome Boole[P] (notare le parentesi quadre, tipiche di *Mathematica*).

Qui ci interessa segnalare in particolare che, usando la notazione di Iverson, la funzione caratteristica di un sottoinsieme A di un insieme E si può scrivere nella forma compatta seguente:

$$\chi(A) = [x \in A]$$

Si potrebbe obiettare che la semplificazione di questa definizione, che è ora fatta "con una linea anziché con due", è stata ottenuta solo a prezzo del trasferimento della "complicazione" più a monte: la situazione sta esattamente in questi termini, ma c'è da osservare che questo trasferimento comporta la possibilità di usare la stessa notazione anche in altri contesti.

Insomma: se si deve fare per forza una certa fatica, conviene farla il più presto possibile, in modo da trarne i massimi vantaggi.

3 Disequazioni in $\mathbb{R}$ e segno delle funzioni

Nella risoluzione di disequazioni nell'insieme dei numeri reali, la determinazione del segno delle funzioni gioca un ruolo fondamentale. In questa sezione consideriamo alcuni esempi significativi.

Esempio 3.1. Supponiamo di dover risolvere le seguenti disequazioni:

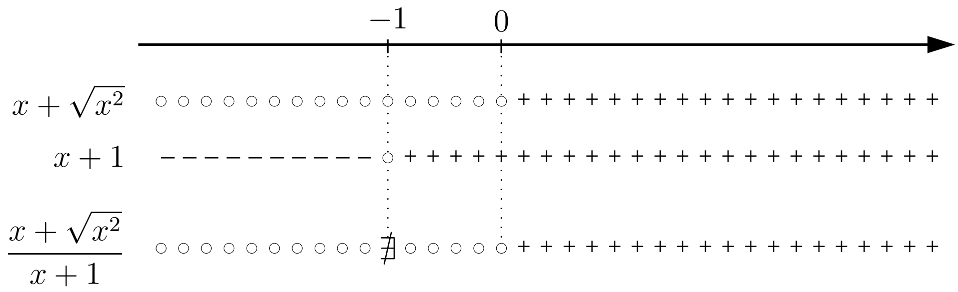
a) $\dfrac{x+\sqrt{x^2}}{x+1} > 0$

b) $\dfrac{x+\sqrt{x^2}}{x+1} < 0$

c) $\dfrac{x+\sqrt{x^2}}{x+1} \geq 0$

d) $\dfrac{x+\sqrt{x^2}}{x+1} \leq 0$

Il metodo più efficiente consiste nel determinare gli insiemi in cui il numeratore e denominatore sono, rispettivamente, maggiori di zero, minori di zero, uguali a zero, e poi fare un grafico, utilizzando opportune notazioni.



Se ne deduce subito che le soluzioni cercate sono, per ognuno dei casi considerati,

a) $]0,+\infty[$

b) $\emptyset$

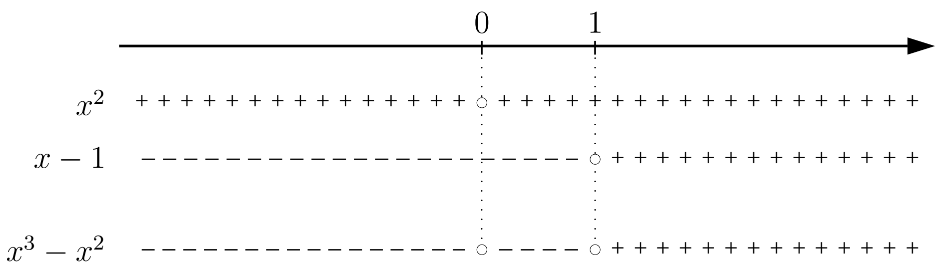
c) $]-\infty,-1[\cup]-1,+\infty[$

d) $]-\infty,-1[\cup]-1,0]$

Esempio 3.2. Trovare il dominio della seguente funzione:

$$f(x)=\sqrt{x^3-x^2}\,.$$

Deve essere $x^3-x^2\geq 0$, ovvero $x^2(x-1)\geq 0$. Procedendo come indicato nell'esempio 3.1, si giunge a costruire il grafico che segue.



A questo punto la conclusione è immediata: il dominio richiesto è $\{0\} \cup [1, +\infty[$.

Esempio 3.3. Trovare il dominio della seguente funzione:

$$f(x) = \sqrt{\frac{\sin x - 1}{2x - \pi}}\,.$$

Deve essere

$$\frac{\sin x - 1}{2x - \pi} \geq 0\,.$$

Si può osservare che $(\sin x - 1)$ è strettamente negativo tranne nei punti $\pi/2 + 2k\pi$, dove vale zero; invece $(2x - \pi)$ è positivo per $x > \pi/2$, negativo per $x < \pi/2$, nullo per per $x = \pi/2$. Si può dunque concludere che il dominio è:

$$\left]-\infty, \frac{\pi}{2}\right[\cup \left\{x = \frac{\pi}{2} + 2k\pi,\ k \text{ intero } \geq 1\right\}\,.$$

Questi esempi rendono palese quanto sia importante nella risoluzione di disequazioni distinguere accuratamente, anche con opportune strategie grafiche, gli insiemi dove le funzioni sono > 0, quelli dove sono < 0, quelli dove sono $= 0$, naturalmente dopo averne trovato il dominio.

4 Continuità, derivata, derivabilità

È opportuno ricordare che una funzione $f\colon A \subseteq \mathbb{R} \to \mathbb{R}$ è *continua* in un punto c del suo dominio se

$$\forall\, U_{f(c)}\ \exists\, U_c \mid \forall\, x \in U_c, f(x) \in U_{f(c)}\,.$$

In altre parole: se c è un punto isolato del dominio di una funzione, la funzione stessa è automaticamente continua in c; se invece c è di accumulazione per il dominio della funzione, la funzione è continua in c se e solo se

$$\lim_{x \to c} f(x) = f(c)$$

Se una funzione non è continua in un punto c del suo dominio, si dice *discontinua* in c. Questo succede se il limite della funzione quando x tende a c non esiste, oppure esiste, ma è diverso dal valore della funzione nel punto c. In sostanza una funzione può essere continua o discontinua *solo* ed *esclusivamente* in un punto c del suo dominio. Purtroppo è invalsa l'abitudine, deprecabile a mio avviso (e naturalmente non solo a mio avviso!), di estendere il concetto di funzione discontinua anche a punti che non appartengono al dominio, ma siano di accumulazione per il dominio stesso. Valga per tutti l'*Osservazione* fatta a questo proposito in Dolcher (1991): "Una funzione che in un punto $\overline{x}$ non sia definita non sarà da dirsi 'discontinua' in tale punto. Ad esempio, funzioni come $1/x$, $x/|x|$, $\operatorname{tg} x$, $\sin 1/x$ sono, come si dimostrerà, continue: nessun punto è di discontinuità per esse."

Ancora più chiara, se possibile, l'*Osservazione* contenuta in Trombetta (2006). "E se x_0 non appartiene al dominio E della funzione? Allora in x_0 f *non è né continua né discontinua*, semplicemente in x_0 la funzione *non c'è*, e ciò anche se x_0 è di accumulazione per E.

Consideriamo la funzione di $\mathbb{R} \setminus \{0\}$ in $\mathbb{R}$ definita da $f(x) = 1/x$. Questa funzione è ovunque continua. Ora il punto 0 non fa parte del dominio di f; non ha dunque senso dire che f è discontinua in 0: la funzione in 0 *non c'è*; il problema della continuità in 0 non si pone. Quello che si può dire è che, in questo caso, f *non* è prolungabile per continuità nel punto 0.

Consideriamo ancora la funzione $f\colon \mathbb{R} \to \mathbb{R}$ definita da $f(x) = {}^{|x|}/_{x}$. Non ammettendo f limite per $x \to 0$, non può essere prolungata per continuità in tale punto. È però ovviamente possibile prolungarla in modo da renderla continua a destra o a sinistra in 0."

Dal punto di vista delle applicazioni ha comunque un notevole interesse la considerazione del comportamento di una funzione nei pressi di un punto di accumulazione del dominio che non appartenga al dominio stesso. Per esempio per la funzione definita da $f(x) = {}^{\sin x}/_{x}$ il punto 0 non fa parte del dominio, dunque non ha senso chiedersi se la funzione sia o meno continua in 0. Si potrà però usare una locuzione speciale per indicare che il comportamento della funzione nei pressi del punto 0 è di grande interesse; una di quelle usate è: 0 è un *punto singolare* per la funzione. Nel seguito mi atterrò a questa convenzione.

Ora è facile provare che tutte le *funzioni elementari* (polinomi, funzioni potenza, funzioni esponenziali, funzioni trigonometriche, e funzioni ottenute mediante loro somme, prodotti, quozienti, composte, inverse) sono continue in *tutti* i punti del loro *dominio naturale*.

Per costruire funzioni non continue con metodi elementari il sistema più semplice è quello di ricorrere alle funzioni caratteristiche. Se per esempio consideriamo la funzione caratteristica dei razionali come sottoinsieme dei reali, funzione che, usando la notazione di Iverson possiamo definire come $f(x) = [x \in \mathbb{Q}]$, otteniamo una funzione discontinua su ogni $x \in \mathbb{R}$. Se invece consideriamo la funzione caratteristica di un sottoinsieme *finito* A di $\mathbb{R}$, otterremo una funzione discontinua esattamente in corrispondenza degli $x \in A$.

Passiamo ora al problema della derivata di una funzione. Data una funzione $f\colon \mathbb{R} \to \mathbb{R}$, e un punto c che questa volta si considera abitualmente *interno* al dominio (ovvero c appartiene all'interno di un intervallo contenuto nel dominio), si considera il *rapporto incrementale* della funzione f, relativo a c:

$$\frac{f(x) - f(c)}{x - c}\,.$$

Si considera poi il limite di questo rapporto incrementale, al tendere di x a c. Si possono presentare le seguenti situazioni:

1. $\displaystyle\lim_{x\to c} \frac{f(x) - f(c)}{x - c}$ non esiste;
2. $\displaystyle\lim_{x\to c} \frac{f(x) - f(c)}{x - c} = \infty$ ($+\infty$ oppure $-\infty$);
3. $\displaystyle\lim_{x\to c} \frac{f(x) - f(c)}{x - c} = l \in \mathbb{R}\,.$

Se per la funzione f si presenta il terzo caso, allora essa si dice *derivabile* nel punto c e l si chiama la *derivata* di f in c (che si indica, per esempio, con $f'(c)$). A questo punto però le convenzioni sulla nomenclatura non sono universali: se tutti concordano, nel terzo caso, nel chiamare l la *derivata* della funzione f nel punto c, per alcuni si parla di derivata anche nel secondo caso, e si usa esplicitamente la nomenclatura *derivata infinita*. Se si fa questa scelta bisogna stare attenti a ricordare, come ben osserva Dolcher (1991), che la parola "derivabile" non sta a significare "che ammette derivata", ma "che ammette derivata finita".

Il motivo di queste scelte è legato alle proprietà geometriche del grafico di una funzione derivabile o che ammette derivata infinita. Precisamente:

1. Se un funzione f è *derivabile* in un punto c, cioè il limite del rapporto incrementale della funzione esiste finito, allora la funzione è *continua* in c e il suo grafico ammette, in $(c, f(c))$, una tangente non verticale.
2. Se una funzione è *continua* in c e il limite del rapporto incrementale è $+\infty$ o $-\infty$, allora il grafico della funzione f ammette, in $(c, f(c))$, una tangente verticale.

Si noti che, mentre nel primo caso la continuità discende dalla derivabilità, nel secondo caso essa deve essere inserita tra le ipotesi. Si possono considerare gli esempi che seguono.

Esempio 4.1. La funzione $f\colon \mathbb{R} \to \mathbb{R}$ data dalla legge $f(x) = \sqrt[3]{x}$ è continua in tutto il suo dominio (come ogni funzione elementare!) ed ha, in 0, tangente verticale.

Esempio 4.2. La funzione $f\colon \mathbb{R}\setminus\{0\} \to \mathbb{R}$ data da $f(x) = {}^{|x|}\!/_{x}$ non è, in zero, né continua né discontinua, in quanto non definita in 0, e dunque non ha senso calcolarne il rapporto incrementale relativo al punto 0. Non c'è modo per prolungare questa funzione in 0 in modo da renderla continua, ma un prolungamento possiamo comunque considerarlo. Possiamo considerare per esempio la funzione

$$g(x) = \begin{cases} {}^{|x|}\!/_{x} & \text{se} \quad x \neq 0 \\ 1 & \text{se} \quad x = 0 \end{cases} .$$

Essa è continua solo da destra in 0 ed è possibile considerare il rapporto incrementale relativo a 0:

$$\frac{\dfrac{|x|}{x} - 1}{x - 0} = \frac{\dfrac{|x|}{x} - 1}{x} .$$

Il limite di questo rapporto incrementale vale 0 da destra e $+\infty$ da sinistra: la funzione ha una derivata destra in 0, mentre non ha derivata sinistra (o ha derivata sinistra infinita, a seconda delle convenzioni).

Avremmo avuto una situazione simmetrica se avessimo scelto di prolungare f in 0 attribuendole il valore -1 anziché 1.

Poiché però la funzione non è prolungabile per continuità, la scelta del valore in 0 è completamente arbitraria. È interessante notare che se scegliamo un valore strettamente compreso tra -1 ed 1 si ottiene una funzione discontinua sia da destra che da sinistra, ma in cui il limite del rapporto incrementale è sempre $+\infty$ (cioè in cui la derivata è infinita, secondo una delle convenzioni citate); se scegliamo un valore strettamente maggiore di 1 si ottiene ovviamente una funzione ancora discontinua sia da destra che da sinistra, ma ora il limite destro del rapporto incrementale è $-\infty$, mentre quello sinistro è $+\infty$; situazione simmetrica se scegliamo un valore strettamente minore di -1.

Per rendersi conto di questi fatti consideriamo le funzioni

$$g_a(x) = \begin{cases} {}^{|x|}\!/_{x} & \text{se} \quad x \neq 0 \\ a & \text{se} \quad x = 0 \end{cases} ,$$

o, per essere più chiari,

$$g_a(x) = \begin{cases} 1 & \text{se} \quad x > 0 \\ a & \text{se} \quad x = 0 \\ -1 & \text{se} \quad x < 0 \end{cases} .$$

Il loro rapporto incrementale relativo al punto 0 vale:

$$\frac{g_a(x) - g_a(0)}{x - 0} = \begin{cases} \frac{1-a}{x} & \text{se} \quad x > 0 \\ \frac{-1-a}{x} & \text{se} \quad x < 0 \end{cases} ,$$

e le conclusioni nei vari casi sono immediate, tenendo conto dei possibili valori di a.

5 Calcoli approssimati

"Il calcolatore automatico a cifre è uno degli ausiliari più potenti dell'intelletto umano, di cui forma per così dire un prolungamento, nello stesso senso che una macchina mossa da un motore forma un prolungamento della muscolatura umana; in quest'ordine di idee, anzi, il calcolatore può essere paragonato solo alle macchine più potenti sinora costruite."

La citazione, (Forsythe, 1973), è presa da un articolo scritto nell'ormai lontanissimo 1969, ma è naturalmente ancora più valida oggi quando i calcolatori possono forse essere considerati semplicemente un'appendice del corpo umano e non più solo del suo intelletto.

I problemi connessi con l'uso dei calcolatori, e la limitazione conseguente alla necessità di implementare solo un sottoinsieme *finito* dell'insieme dei razionali, sono ben noti e portano a situazioni sorprendenti. Si può considerare, come utile esempio, il problema, preso dal citato articolo di Forsythe (1973), della risoluzione della equazione di secondo grado

$$10^{-30}x^2 + 10^{30}x + 10^{30} = 0\,.$$

Se si prova a risolverlo con *Derive* si ottiene un risultato completamente diverso a seconda del numero di cifre decimali esatte che vengono utilizzate. Precisamente, fino a 58 cifre decimali si ottiene $x_1 = 1.5$, $x_2 \approx 10^{60}$, da 59 cifre decimali in poi si ottiene invece $x_1 = 1$, $x_2 \approx 10^{60}$. Come si vede una delle due radici rimane stabile indipendentemente dal numero di cifre decimali usate, mentre l'altra no (e la differenza tra i risultati, seppur piccola in termini assoluti, non è trascurabile in termini relativi in quanto ammonta al 50%) Risultati diversi si ottengono con *Mathematica*, che utilizza diversi algoritmi di calcolo interno, estremamente efficienti.

Il problema diventa particolarmente interessante con la funzione signum. Questa volta usiamo *Mathematica*, per evitare sorprese. Ricordo che, in *Mathematica*, $\operatorname{sgn}(0) = 0$, e che la funzione signum è denotata con `Sign`. Riporto di seguito i risultati di alcuni calcoli, relativi a

$$\operatorname{sgn}\left(\sqrt{2} - \frac{2}{\sqrt{2}}\right) \quad \text{oppure} \quad \operatorname{sgn}\left(-\sqrt{2} + \frac{2}{\sqrt{2}}\right).$$

1. `Sign[Sqrt[2] - 2/Sqrt[2]] = 0`
2. `Sign[-Sqrt[2] + 2/Sqrt[2]] = 0`
3. `Sign[Sqrt[2.0] - 2.0/Sqrt[2.0]] = 1`
4. `Sign[-Sqrt[2.0] + 2.0/Sqrt[2.0]] = -1`
5. `Sign[Sqrt[2.000000000000000] -`
   `2.000000000000000/Sqrt[2.000000000000000]] = 1`
6. `Sign[Sqrt[2.0000000000000000] -`
   `2.0000000000000000/Sqrt[2.0000000000000000]] = 0`

I risultati sono, a prima vista, sorprendenti, in quanto se si conviene che $\operatorname{sgn}(0) = 0$, si dovrebbe ottenere 0 in tutti i casi. La situazione si chiarisce se si osserva che mentre nei primi due casi il software ha usato le tecniche del calcolo simbolico (come avremmo fatto noi "a mano"), negli altri casi ha usato le tecniche dell'analisi numerica, con un opportuno numero di cifre decimali. In particolare la differenza tra il caso n.5 e il caso n.6 è legata al passaggio da 16 a 17 cifre decimali esatte.

Questi esempi chiariscono anche il perchè un software come *Derive* usa la convenzione $\operatorname{sgn}(0) = \pm 1$: poichè il risultato del calcolo del segno di numeri "prossimi a zero" dipende dal numero di cifre decimali usate, tanto vale avvertire il programmatore, il quale saprà poi prendere le decisioni giuste in base al contesto o ad altre considerazioni che il software non può conoscere.

Da un punto di vista formale, però, si tratta di una convenzione pericolosa, perchè introduce, di soppiatto, l'uso di funzioni plurivoche che, in particolare nei sistemi di calcolo automatico, possono produrre gravi problemi di indecisione. Per esempio, sempre usando *Derive*, il calcolo di "sgn(0) + 2 sgn(0)" produce come risultato un significativo " ? ", in quanto il software non riesce a fare alcuna scelta.

Problemi come quello qui proposto hanno grande rilevanza nell'analisi numerica: l'implementazione in un algoritmo di calcolo di funzioni che presentino discontinuità o singolarità è un problema spesso considerato *mal posto*, e richiede grande attenzione. Con riferimento all'esempio sopra discusso, il problema è che, da un punto di vista puramente matematico 0 è 0 e basta, da un punto di vista numerico tutto dipende da come lo 0 viene calcolato: calcolare sgn(0) è completamente diverso dal calcolare sgn(*expr*), dove *expr* è un'espressione che darebbe 0 in un calcolo "esatto", ma che può dare "quasi 0" in un calcolo numerico.

Per rendersi conto del problema si provi a ripetere i calcoli precedenti, sempre con *Mathematica*, ma questa volta usando la funzione $f(x) = 1/x$. Si ottengono i seguenti risultati:

1. `1/(Sqrt[2] - 2/Sqrt[2])` : Infinite expression $1/0$ encountered;
2. `1/(Sqrt[2.0] - 2.0/Sqrt[2.0])` $= 4.5036 \times 10^{15}$;
3. `1/(-Sqrt[2.0] + 2.0/Sqrt[2.0])` $= -4.5036 \times 10^{15}$;

e questa volta la differenza tra il secondo e il terzo risultato è veramente abissale.

6 La funzione signum

Per ogni numero reale x, la funzione *valore assoluto* fornisce la distanza del numero stesso dall'origine, nella usuale metrica dei reali. Questa funzione può essere definita in vario modo. La definizione classica è:

$$\mathrm{abs}(x) = |x| = \begin{cases} x & \text{se} \quad x \geq 0 \\ -x & \text{se} \quad x < 0 \end{cases} .$$

Altre definizioni possibili sono le seguenti:

$$\mathrm{abs}(x) = |x| = \sqrt{x^2} \quad \text{oppure} \quad \mathrm{abs}(x) = |x| = \max(x, -x) .$$

In sostanza la funzione valore assoluto fornisce una misura di *quanto grande* sia un numero reale, indipendentemente dal fatto di essere positivo o negativo, ovvero *estrae* da un numero reale la sua *grandezza assoluta*.

È allora naturale pensare di costruire anche una funzione che estragga da ogni numero reale il segno, in modo da poter poi scrivere il numero stesso come combinazione (prodotto) del suo segno e del suo valore assoluto. Purtroppo, mentre l'introduzione del valore assoluto porta naturalmente ad una funzione continua e costruibile anche a partire dalle funzioni elementari ($|x| = \sqrt{x^2}$), ciò non avviene per la funzione segno.

È immediato che se consideriamo una funzione $f\colon \mathbb{R} \to \mathbb{R}$ tale che $f(x)$ valga 1 sui reali positivi, -1 sui reali negativi, e che assuma un qualunque valore a sullo zero, sarà possibile scrivere ogni numero reale nella forma $x = f(x) \cdot |x|$.

Si potrebbe anche evitare di definire la funzione nello 0 (in fondo 0 non appartiene né al gruppo abeliano moltiplicativo dei reali positivi, né al sottoinsieme dei reali negativi). A me pare che questa scelta sia illogica proprio in vista dello scopo principale per cui questa funzione viene introdotta, e cioè quello di scrivere *ogni* numero reale come prodotto del suo segno e del suo valore assoluto. Quanto poi alla scelta di come definirla in 0 (scelta arbitraria, come detto), mi pare che la più coerente sia quella di porla uguale a 0: se ne mantiene in questo modo la caratteristica essenziale, e cioè quella di essere dispari, ovvero simmetrica rispetto all'origine.

Ritengo pertanto che la scelta più opportuna per definire la funzione signum sia quella adottata dalla quasi totalità dei software di calcolo e dei linguaggi di programmazione:

$$\operatorname{sgn}(x) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{se} \quad x > 0 \\ 0 & \text{se} \quad x = 0 \\ -1 & \text{se} \quad x < 0 \end{cases} .$$

Con ciò la formula

$$x = \operatorname{sgn}(x) \cdot |x|$$

è valida per *ogni* numero reale x e, inoltre, la funzione sgn è simmetrica rispetto all'origine.

Se poi si usano le notazioni di Iverson, allora la simmetria della definizione che ho appena proposto diventa evidente:

$$\operatorname{sgn}(x) = [x > 0] - [x < 0] = [x \geq 0] - [x \leq 0] .$$

In sostanza la funzione signum, come appena definita, è la differenza tra la funzione caratteristica dei reali strettamente positivi e quella dei reali strettamente negativi, o anche tra la funzione caratteristica dei reali positivi o nulli e quella dei reali negativi o nulli.

Molti altri sono i vantaggi della scelta fatta. Ne elenco alcuni.

1. Se non volessi definire la funzione signum in 0 basterebbe considerare la funzione $|x|/x$, senza bisogno di introdurre nuove terminologie e nuovi simboli. È questa, per esempio, la scelta fatta in Prodi (1970), dove si introduce la funzione segno come $|x|/x$, ma senza definire alcun simbolo specifico.
2. La definizione della funzione *anche* in 0 consente di considerare un esempio semplice di funzione discontinua: se non la definissi in 0 sarebbe sempre continua nel suo dominio, mentre in 0 non sarebbe né continua né discontinua, come già osservato nel §4, ma semplicemente inesistente. A livello didattico è molto utile avere esempi semplici di funzioni discontinue, in quanto *tutte* le funzioni elementari sono invece continue nel loro dominio naturale.
3. Attribuire alla funzione, in 0, un valore strettamente compreso tra -1 e 1 consente come già osservato a pag.8, di avere un esempio di funzione che ha limite del rapporto incrementale uguale a $+\infty$, senza essere continua nel punto. Questa è una situazione didatticamente efficace perchè mostra quanto sia importante nella dimostrazione della continuità delle funzioni derivabili il fatto che il limite del rapporto incrementale sia *finito*.

C'è poi da considerare lo stretto legame che si viene a creare tra questa funzione e la funzione di *Heaviside* o *funzione gradino unitaria*, abitualmente indicata con H, e definita in vari modi mediante un limite. Una delle definizioni più comuni, e più semplici, è:

$$\mathrm{H}(x) \stackrel{\text{def}}{=} \lim_{t \to 0} \frac{1}{1 + \mathrm{e}^{-x/t}} ,$$

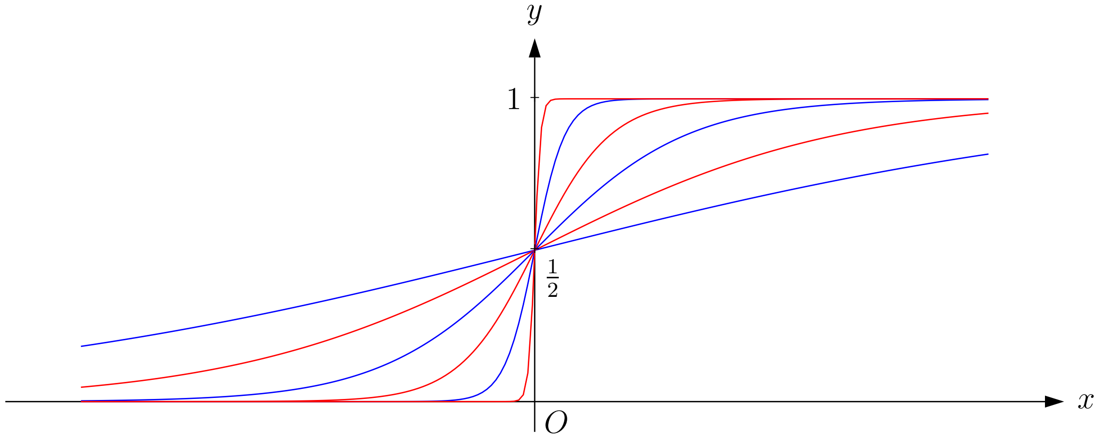
che produce il risultato

$$\mathrm{H}(x) = \begin{cases} 1 & \text{se} \quad x > 0 \\ 1/2 & \text{se} \quad x = 0 \\ 0 & \text{se} \quad x < 0 \end{cases} .$$

Nel grafico che segue si possono visualizzare le funzioni

$$f_t(x) = \frac{1}{1 + \mathrm{e}^{-x/t}}$$

per alcuni valori di t, per rendersi conto di come, al limite, si ottenga proprio la funzione *gradino unitaria*. I valori di t considerati sono, nell'ordine, 2, 1, 0.5, 0.25, 0.02.



Con la definizione data di funzione signum, il legame con la funzione di Heaviside diventa:

$$\operatorname{sgn}(x) = 2\mathrm{H}(x) - 1 .$$

La funzione di Heaviside è importante nelle applicazioni. Può essere usata per esempio come modello matematico in un circuito in cui viene fatta improvvisamente circolare corrente ad un certo istante. Come già per la funzione signum, dal punto di vista applicativo il suo valore in zero è poco importante, ma la simmetria che si ottiene e la possibilità di definirla come limite di funzioni continue, induce ad attribuirle il valore $1/2$ nell'origine.

Naturalmente anche la funzione signum può essere definita mediante limite di funzioni continue, e la cosa è di grande interesse, perchè significa che le funzioni che al limite producono la nostra funzione possono essere usate come sue approssimazioni continue. Une delle possibili definizioni è:

$$\operatorname{sgn}(x) = \lim_{k \to +\infty} \operatorname{tgh}(kx) ,$$

ove tgh è la funzione tangente iperbolica.

Ma il motivo che mi ha spinto a parlare della funzione di Heaviside è legato ad un fatto di particolare importanza nella teoria delle distribuzioni. Anche se qui non è possibile addentrarsi nei dettagli di questa teoria, è notevole il fatto che

$$\frac{\mathrm{dH}(x)}{\mathrm{d}x} = \delta(x) ,$$

dove δ è la distribuzione *delta di Dirac*.

La distribuzione *delta di Dirac* è spesso impropriamente chiamata funzione e se ne usano le proprietà formali seguenti

$$\delta(x)=\begin{cases} +\infty & \text{se} \quad x=0 \\ 0 & \text{se} \quad x\neq 0 \end{cases}, \quad \int_{-\infty}^{+\infty}\delta(x)\,\mathrm{d}x=1\,, \quad \int_{-\infty}^{+\infty}f(x)\delta(x)\,\mathrm{d}x=f(0)\,,$$

in maniera un po' "allegra" e senza preoccuparsi troppo della precisione. Per fortuna che (quasi) tutto funziona bene...! Queste proprietà estendono al continuo le analoghe proprietà del simbolo *Delta di Kronecker*.

La citata proprietà della funzione di Heaviside si traduce nella seguente proprietà della funzione signum:

$$\frac{\mathrm{d}\,\mathrm{sgn}(x)}{\mathrm{d}x}=2\,\delta(x)\,.$$

7 Funzione signum e funzione valore assoluto

La funzione valore assoluto, come è ben noto, non è derivabile nello 0, dove ha un *punto angoloso* con derivata sinistra -1 e destra $+1$. Al di fuori dello 0 si ha invece coincidenza tra la derivata della funzione valore assoluto e la funzione signum:

$$\forall\, x\in\mathbb{R}\setminus\{0\} \quad , \quad \frac{\mathrm{d}\big(\mathrm{abs}(x)\big)}{\mathrm{d}x}=\mathrm{sgn}(x)\,.$$

La funzione signum è invece una funzione integrabile, secondo Riemann, in qualunque intervallo di $\mathbb{R}$. Ha dunque senso calcolarne la *funzione integrale*, di punto iniziale un punto c qualunque. Se si considera in particolare la funzione integrale di punto iniziale 0 si ha la seguente, interessante proprietà:

$$\mathrm{abs}(x)=|x|=\int_0^x \mathrm{sgn}(t)\,\mathrm{d}t\,,$$

valida per tutti i reale x.

In base alle note proprietà delle funzioni integrali si possono dedurre le ben note proprietà della funzione valore assoluto, in particolare il fatto che essa è continua in tutto $\mathbb{R}$ e derivabile per $x\neq 0$, mentre in 0 ha derivate sinistra e destra diverse ed uguali, rispettivamente, a -1 e 1.

Ricordo che, per una nota proprietà dell'integrale di Riemann, il valore della funzione signum non ha alcun interesse per il calcolo del precedente integrale: basta che la funzione stessa abbia, in 0, un valore qualunque.

La precedente formula potrebbe anche essere usata come *definizione della funzione* valore assoluto. Per completezza riporto qui le varie definizioni che sono state date di questa funzione:

$$\begin{aligned}\mathrm{abs}(x)=|x| &= \begin{cases} x & \text{se} \quad x\geq 0 \\ -x & \text{se} \quad x<0 \end{cases} \\ &= \sqrt{x^2} \\ &= \max(x,-x) \\ &= \int_0^x \mathrm{sgn}(t)\,\mathrm{d}t\,.\end{aligned}$$

8 Qualche interessante esempio di funzione

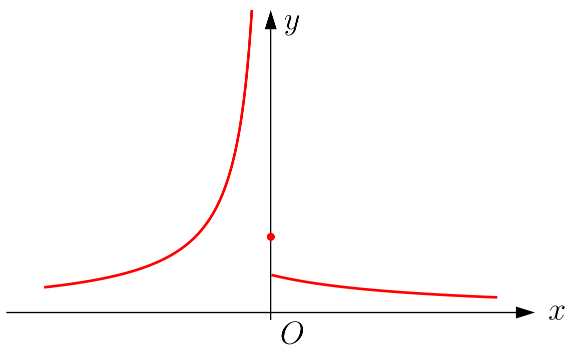
Utilizzando la funzione signum definita nel §6 si possono costruire parecchie funzioni interessanti e utili per produrre esempi significativi nella teoria delle funzioni reali di variabile reale.

Ne propongo qui un paio.

Esempio 8.1. La funzione $f\colon \mathbb{R} \to \mathbb{R}$ definita da,

$$f(x) = \frac{1}{(x+1)\operatorname{sgn}(x)+1},$$

ha l'interessante proprietà di avere un punto in comune, e precisamente il punto (0,1), col suo unico asintoto verticale che è l'asse delle ordinate. Nessuna funzione elementare interseca i suoi eventuali asintoti verticali, per cui questa è una proprietà notevole. È chiaro che una funzione con questa proprietà avrebbe anche potuto essere costruita ad hoc, ma la cosa interessante è che è possibile costruirla usando la funzione signum. Il grafico di questa funzione è riportato di seguito.



Esempio 8.2. La funzione $f\colon \mathbb{R}\setminus\{0\} \to \mathbb{R}$, definita da

$$f(x) = \left|\operatorname{sgn}\left(x\sin\frac{1}{x}\right)\right|,$$

è molto interessante come esempio di non applicabilità del "cambiamento di variabile" nel calcolo dei limiti. Infatti, tenendo conto che

$$\lim_{x\to 0} x\sin\frac{1}{x} = 0\,, \qquad \text{e che} \qquad \lim_{t\to 0} |\operatorname{sgn}(t)| = 1$$

si potrebbe essere portati ad operare come di seguito indicato.

Nel calcolo del

$$\lim_{x\to 0}\left|\operatorname{sgn}\left(x\sin\frac{1}{x}\right)\right| = 0\,,$$

se "sostituisco"

$$x\sin\frac{1}{x} \quad \text{con} \quad t\,,$$

sono ricondotto al calcolo di

$$\lim_{t\to 0} |\operatorname{sgn}(t)|\,,$$

che sappiamo essere 1. Dunque posso concludere che il limite richiesto vale 1. Il ragionamento è errato, in quanto la funzione f assume infinite volte il valore 0 in ogni intorno dell'origine e quindi non può avere limite 1.

Il teorema sul limite delle funzioni composte non è molto facile. Una versione semplificata, che fornisce una condizione sufficiente per l'esistenza, è la seguente:
Se f e g sono due funzioni tali che abbia senso considerare la funzione composta e i limiti che seguono, e se $\lim_{x\to c} f(x) = l$ e $\lim_{t\to l} g(t) = m$, con la condizione che $f(x) \neq l$ almeno per $x \neq c$ allora

$$\lim_{x\to c} f(g(x)) = m\,.$$

Se poi g è continua in l, oppure non è definita in l, la condizione $f(x) \neq l$ è superflua.

Anche se normalmente un teorema come questo non viene proposto nei corsi base di analisi, credo sia comunque opportuno fornire almeno un esempio delle difficoltà che si possono incontrare nell'applicare la tecnica di "sostituzione" per il calcolo dei limiti: non si tratta dello stesso tipo di procedimento che si usa quando, per esempio, per risolvere un'equazione biquadratica si sostituisce x^2 con t.

9 Un'applicazione interessante

L'applicazione qui di seguito proposta, didatticamente molto interessante, è presa dal sito web dell'Unione Matematica Italiana, e precisamente da http://umi.dm.unibo.it/italiano/Matematica2003/seconda/Mat_2010.pdf. Rimando a questo articolo per ogni utile approfondimento.

Si tratta di costruire un algoritmo automatico che permetta di decidere se un punto P di un piano sia *dentro*, *fuori*, oppure *sul bordo* di un dato triangolo del piano. L'algoritmo proposto utilizza la funzione segno, naturalmente con 3 valori come fatto nel §6, visto che si devono valutare 3 situazioni diverse.

L'idea di base è facile da descrivere. Una retta nel piano cartesiano è data da un'equazione di primo grado in due incognite: $ax + by + c = 0$. Se un punto $P(x_0, y_0)$ appartiene alla retta, allora il polinomio $ax + by + c$ si annulla in corrispondenza di (x_0, y_0); se invece P non appartiene alla retta, allora lo stesso polinomio assume valori positivi o negativi a seconda che P stesso si trovi in uno o nell'altro dei due semipiani individuati dalla retta.

Dati 3 punti, M, N, P, del piano cartesiano si costruisce innanzitutto una funzione, $\operatorname{sgn}(M, N, P)$, che restituisce il valore 1 se i tre punti M, N, P si succedono, nel triangolo di cui sono vertici, in senso antiorario, -1 se in senso orario, 0 se il triangolo è degenere. Si osserva poi che due punti P e Q appartengono allo stesso semipiano rispetto alla retta individuata da M ed N se e solo se la funzione segno prima definita fornisce lo stesso valore.

Ora, dato un triangolo $\triangle ABC$ e un punto P nel piano, si considerano successivamente le tre rette AB, BC, AC. Per ciascuna di esse si controlla, mediante la funzione $\operatorname{sgn}(M, N, P)$, come si dispone il punto P rispetto al restante vertice del triangolo. Se almeno uno dei tre valori è -1, il punto P è esterno al triangolo, se tutti tre i valori sono 1, il punto P è interno al triangolo, se nessun valore è negativo e almeno uno è 0, allora il punto P sta sul triangolo.

Riferimenti bibliografici

Borowski, Ephraim J. e Borwein, Jonathan M. (2002). *Dictionary of Mathematics*. Collins, Glasgow, 2ª edizione.

Dolcher, Mario (1991). *Elementi di analisi matematica*. Lint, Trieste.

Forsythe, George E. (1973). Come risolvere un'equazione quadratica su un calcolatore. *Le scienze matematiche*, pp. 158–176. Raccolta di saggi a cura dell'Unione matematica Italiana, edito da Zanichelli, Bologna. Traduzione di Fernando Bortolini da *The Mathematical Sciences*, Mass. Inst. of Technology, 1969.

Gelbaum, Bernard R. e Olmsted, John M.H. (1979). *Controesempi in analisi matematica*. Mursia, Milano. Traduzione di Pietro e Cesira Canetta.

Knuth, Donald E. (1992). Two notes on notation. *American Mathematical Monthly*, **99**(5), 403–422.

Laforgia, Andrea (1995). *Precalculus*. Cedam, Padova.

Lesina, Roberto (1994). *Il nuovo manuale di stile*. Zanichelli, Bologna, 2ª edizione.

Prodi, Giovanni (1970). *Analisi Matematica*. Borignhieri, Torino, 1ª edizione.

Trombetta, Maurizio (2006). *Appunti del corso di Analisi Matematica per il Diploma universitario*. Dispense. Materiale reperibile in http://www.dmi.units.it/~obersnel/.