

On an Adaptive Regularization for Ill-posed Nonlinear Systems and its Trust-Region Implementation

Elisa Riccietti

Università degli Studi di Firenze

Dipartimento di Matematica e Informatica 'Ulisse Dini'

Lavoro in collaborazione con Stefania Bellavia, Benedetta Morini



UNIVERSITÀ
DEGLI STUDI
FIRENZE

XX Congresso dell'Unione Matematica Italiana, Siena, 7-12 Settembre
2015.

Problemi mal posti

Consideriamo il seguente **problema inverso**: dati $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ non lineare, differenziabile con continuità e $y \in \mathbb{R}^n$, determinare x tale che

$$F(x) = y.$$

Definizione

Il problema si dice **ben posto secondo Hadamard** se:

- 1 $\forall y \in \mathbb{R}^n \exists x \in \mathbb{R}^n$ tale che $F(x) = y$ (esistenza),
- 2 F è iniettiva (unicità),
- 3 F^{-1} è continua (stabilità).

Viceversa, il problema si dice **mal posto** quando non è verificata una delle tre proprietà .

Problemi mal posti

- Consideriamo problemi della forma $F(x) = y$ in $(\mathbb{R}^n, \|\cdot\|_2)$, derivanti dalla discretizzazione di un sistema che modella un **problema mal posto**, tali che:
 - la soluzione non sia unica,
 - non valga la proprietà di stabilità .
- In una situazione realistica i **dati del problema y sono perturbati da rumore**, sarà noto quindi solo y^δ tale che:

$$\|y - y^\delta\| \leq \delta$$

con δ soglia del rumore.

- Il problema da considerare in pratica sarà quindi un sistema perturbato della forma:

$$F(x) = y^\delta.$$

Necessità di regolarizzare

- Poichè non vale la proprietà di stabilità le soluzioni del problema originale non dipendono con continuità dai dati.
⇒ Le soluzioni del problema perturbato potrebbero non essere significative come approssimazioni di soluzioni del problema originale.

Necessità di regolarizzare

- Poichè non vale la proprietà di stabilità le soluzioni del problema originale non dipendono con continuità dai dati.
⇒ Le soluzioni del problema perturbato potrebbero non essere significative come approssimazioni di soluzioni del problema originale.
- In questo contesto non è possibile assumere la limitatezza della norma dell'inversa della Jacobiana di F , come nel caso dei problemi ben posti.

Necessità di regolarizzare

- Poichè non vale la proprietà di stabilità le soluzioni del problema originale non dipendono con continuità dai dati.
⇒ Le soluzioni del problema perturbato potrebbero non essere significative come approssimazioni di soluzioni del problema originale.
- In questo contesto non è possibile assumere la limitatezza della norma dell'inversa della Jacobiana di F , come nel caso dei problemi ben posti.
- I metodi usati per problemi ben posti non sono adatti in questo contesto.



Necessità di regolarizzare

Argomenti della presentazione

- Introduzione ai metodi di regolarizzazione iterativi.
- Descrizione dei metodi Levenberg-Marquardt e loro variante regolarizzante.
- Introduzione di una nuova variante regolarizzante dei metodi trust-region, ottenuta con un'opportuna scelta del raggio di trust-region.
- Descrizione delle sue proprietà regolarizzanti e di convergenza.
- Sperimentazione numerica: confronto del metodo proposto con metodi Levenberg-Marquardt regolarizzanti e trust-region classica.

Metodi di Regolarizzazione iterativi

Ipotesi: assumiamo che esista una soluzione x^\dagger di $F(x) = y$.

I **metodi di regolarizzazione iterativi** producono una successione $\{x_k^\delta\}$ tale che se il processo viene arrestato all'indice $k^*(\delta)$ allora:

- $x_{k^*(\delta)}^\delta$ è una approssimazione di x^\dagger ;
- $\{x_{k^*(\delta)}^\delta\}$ tende a x^\dagger quando δ tende a zero;
- nel caso di dati esatti il metodo garantisce convergenza locale a x^\dagger .

Metodi esistenti in letteratura

- Landweber (metodi di tipo gradiente)[Hanke, Neubauer, Scherzer, 1995,Kaltenbacher, Neubauer, Scherzer, 2008]
- Newton troncato - gradiente coniugato [Hanke,1997, Rieder, 2005]
- Metodi Gauss-Newton iterativi regolarizzati [Bakushinsky, 1992, Blaschke, Neubauer, Scherzer, 1997]
- **Levenberg-Marquardt** [Hanke,1997,2010,Vogel 1990, Kaltenbacher, Neubauer, Scherzer, 2008]

L'analisi di questi metodi è esclusivamente locale, **la definizione di metodi globalmente convergenti è un problema ancora aperto.**

Metodo Levenberg-Marquardt

- Dato $x_k^\delta \in \mathbb{R}^n$ e $\lambda_k > 0$, detta J la Jacobiana di F , il passo p_k è il minimizzatore di

$$m_k^{LM}(p) = \frac{1}{2} \|F(x_k^\delta) - y^\delta + J(x_k^\delta)p\|^2 + \frac{1}{2} \lambda_k \|p\|^2;$$

- p_k è la soluzione di

$$(B_k + \lambda_k I)p_k = -g_k$$

con $B_k = J(x_k^\delta)^T J(x_k^\delta)$, $g_k = J(x_k^\delta)^T (F(x_k^\delta) - y^\delta)$;

- il passo viene usato per formare la nuova iterata

$$x_{k+1}^\delta = x_k^\delta + p_k.$$

Metodo Levenberg-Marquardt regolarizzante

- Il parametro $\lambda_k > 0$ viene scelto in modo che:

$$\|F(x_k^\delta) - y^\delta + J(x_k^\delta)p\| = q \|F(x_k^\delta) - y^\delta\|$$

con $q \in (0, 1)$ costante fissata;

- Se i dati sono affetti da rumore il procedimento viene arrestato all' iterazione $k^*(\delta)$ tale che $x_{k^*(\delta)}^\delta$ soddisfa il **principio di discrepanza**:

$$\|F(x_{k^*(\delta)}^\delta) - y^\delta\| \leq \tau \delta < \|F(x_k^\delta) - y^\delta\|$$

per $0 \leq k < k^*(\delta)$ e $\tau > 1$ costante opportunamente scelta.

[Hanke, 1997,2010]

Ipotesi per l'analisi del metodo

L'analisi del metodo viene fatta sotto le seguenti ipotesi:

Data l' approssimazione iniziale x_0 , esistono costanti positive ρ e c tali che

- il sistema $F(x) = y$ sia risolubile in $B_\rho(x_0)$;
- per $x, \tilde{x} \in B_{2\rho}(x_0)$

$$\|F(x) - F(\tilde{x}) - J(x)(x - \tilde{x})\| \leq c\|x - \tilde{x}\|\|F(x) - F(\tilde{x})\|.$$

[Hanke, 1997,2010]

Data la mal posizione del problema non possiamo assumere che l'inversa della Jacobiana di F abbia norma limitata.

Proprietà regolarizzanti del Levenberg-Marquardt

Scegliendo λ_k come soluzione di

$$\|F(x_k^\delta) - y^\delta + J(x_k^\delta)p\| = q\|F(x_k^\delta) - y^\delta\|$$

e arrestando il procedimento quando il principio di discrepanza

$$\|F(x_{k^*(\delta)}^\delta) - y^\delta\| \leq \tau\delta < \|F(x_k^\delta) - y^\delta\|$$

è soddisfatto, si ottiene:

- Nel caso di **dati esatti** ($\delta = 0$): convergenza locale ad x^\dagger ,
- Nel caso di **dati affetti da rumore** ($\delta > 0$): se $\tau > \frac{1}{q}$ per x_0 sufficientemente vicino a x^\dagger il principio di discrepanza è soddisfatto dopo un numero finito $k^*(\delta)$ di iterazioni e $\{x_{k^*(\delta)}^\delta\}$ converge ad una soluzione di $F(x) = y$ se δ tende a zero.

Il metodo è regolarizzante

Metodi trust-region

- Dato $x_k^\delta \in \mathbb{R}^n$, il passo p_k è il minimizzatore di

$$\min_p m_k^{TR}(p) = \frac{1}{2} \|F(x_k^\delta) - y^\delta + J(x_k^\delta)p\|^2,$$

$$\text{t.c. } \|p\| \leq \Delta_k,$$

con $\Delta_k > 0$ raggio della trust-region.

- Posta $\Phi(x) = \frac{1}{2} \|F(x) - y^\delta\|^2$, si calcola

$$\pi_k(p_k) = \frac{\Phi(x_k) - \Phi(x_k + p_k)}{m_k^{TR}(0) - m_k^{TR}(p_k)}.$$

- Fissato $\eta \in (0, 1)$:
 - Se $\pi_k < \eta$ allora $\Delta_{k+1} < \Delta_k$ e $x_{k+1} = x_k$.
 - Se $\pi_k \geq \eta$ allora $\Delta_{k+1} \geq \Delta_k$ e $x_{k+1} = x_k + p_k$.

Metodi trust-region

Si può dimostrare che p_k risolve

$$(B_k + \lambda_k I)p_k = -g_k$$

per qualche $\lambda_k \geq 0$ tale che

$$\lambda_k(\|p_k\| - \Delta_k) = 0,$$

dove abbiamo posto $B_k = J(x_k^\delta)^T J(x_k^\delta)$ e $g_k = J(x_k^\delta)^T (F(x_k^\delta) - y^\delta)$.

Metodi trust-region

Da $\lambda_k(\|p_k\| - \Delta_k) = 0$ segue che:

- Se la soluzione di minima norma p^* di $B_k p = -g_k$ soddisfa $\|p^*\| \leq \Delta_k$ allora $\lambda_k = 0$ e $p_k = p(0)$;
- altrimenti $\lambda_k \neq 0$, $\|p_k\| = \Delta_k$ e $p_k = p(\lambda_k)$ è un passo di Levenberg-Marquardt.



- La trust-region classica non garantisce proprietà regolarizzanti.
- Per avere un metodo regolarizzante la trust-region deve essere attiva:

$$\|p_k\| = \Delta_k.$$

Una versione regolarizzante della trust-region

- Poichè i passi dei metodi Levenberg-Marquardt e trust-region hanno la stessa forma, esiste un forte legame tra i due metodi.
- In analogia a quanto fatto da Hanke **possiamo introdurre una variante dei metodi trust-region che sia regolarizzante** e in più risulti globalmente convergente?

Obiettivi

La trust-region classica è modificata in modo che:

- sia garantita la riduzione monotona della funzione

$$\Phi(x) = \frac{1}{2} \|F(x) - y^\delta\|^2,$$

- sia verificata la **q-condition**:

$$\|F(x_k^\delta) - y^\delta + J(x_k^\delta)p\| \geq q \|F(x_k^\delta) - y^\delta\|.$$

La q-condition è una versione rilassata di

$$\|F(x_k^\delta) - y^\delta + J(x_k^\delta)p\| = q \|F(x_k^\delta) - y^\delta\|.$$

Trust-region regolarizzante

Mostriamo la nuova variante dei metodi trust-region che grazie ad un opportuno aggiornamento della trust-region garantisce:

- il soddisfacimento della q -condition,
- le stesse proprietà di regolarizzazione del metodo Levenberg-Marquardt regolarizzante.

Scelta del raggio di trust-region

Lemma

Sia p_k soluzione del problema di trust-region. Se

$$\Delta_k \leq \frac{1-q}{\|B_k\|} \|g_k\|$$

allora p_k soddisfa la q -condition.

Conseguenza: scelta Δ_k

$$\Delta_k \in \left[C_{\min} \|g_k\|, \min \left\{ C_{\max}, \frac{1-q}{\|B_k\|} \|g_k\| \right\} \right],$$

con C_{\min}, C_{\max} costanti opportune, $B_k = J(x_k^\delta)^T J(x_k^\delta)$ e
 $g_k = J(x_k^\delta)^T (F(x_k^\delta) - y^\delta)$.

Algoritmo : k -esima iterazione della trust-region regolarizzante

Dati x_k^δ , $\eta \in (0, 1)$, $\gamma \in (0, 1)$, $0 < C_{\min} < C_{\max}$.

Dati esatti: dati y , $q \in (0, 1)$.

Dati perturbati: dati y^δ , $q \in (0, 1)$, $\tau > 1/q$.

1. Calcolare $B_k = J(x_k^\delta)^T J(x_k^\delta)$ e $g_k = J(x_k^\delta)^T (F(x_k^\delta) - y^\delta)$.

2. Scegliere $\Delta_k \in \left[C_{\min} \|g_k\|, \min \left\{ C_{\max}, \frac{1-q}{\|B_k\|} \right\} \|g_k\| \right]$

3. Ripeti

3.1 Calcola la soluzione p_k del problema di trust-region.

3.2 Calcola

$$\pi_k(p_k) = \frac{\Phi(x_k^\delta) - \Phi(x_k^\delta + p_k)}{m_k^{TR}(0) - m_k^{TR}(p_k)}$$

con $\Phi(x) = \frac{1}{2} \|F(x) - y^\delta\|^2$, $m_k^{TR}(p) = \frac{1}{2} \|F(x_k^\delta) + J(x_k^\delta)p\|^2$.

3.3 Se $\pi_k(p_k) < \eta$, poni $\Delta_k = \gamma \Delta_k$.

Fino a che $\pi_k(p_k) \geq \eta$.

4. Poni $x_{k+1}^\delta = x_k^\delta + p_k$.

Ipotesi per analisi locale del metodo

Ipotesi 1: Le stesse assunte per il metodo Levenberg-Marquardt. Assumiamo che per un indice \bar{k} esistano costanti positive ρ e c tali che

1 il sistema $F(x) = y$ sia risolubile in $B_\rho(x_k^\delta)$;

2 per $x, \tilde{x} \in B_{2\rho}(x_k^\delta)$

$$\|F(x) - F(\tilde{x}) - J(x)(x - \tilde{x})\| \leq c\|x - \tilde{x}\|\|F(x) - F(\tilde{x})\|.$$

Ipotesi 2: Esiste una costante positiva K_J tale che

$$\|J(x)\| \leq K_J$$

per ogni $x \in \mathcal{L} = \{x \in \mathbb{R}^n \text{ t.c. } \Phi(x) \leq \Phi(x_0)\}$.

Risultati per $\delta = 0$

Lemma

Il metodo descritto genera una successione $\{x_k\}$ tale che per $k \geq \bar{k}$

- la trust-region è attiva, cioè $\lambda_k > 0$;
- x_k appartiene a $B_{2\rho}(x_{\bar{k}})$ e a $B_\rho(x^\dagger)$;
- $\|x_{k+1} - x^\dagger\| < \|x_k - x^\dagger\|$;
- esiste una costante $\bar{\lambda} > 0$ tale che $\lambda_k \leq \bar{\lambda}$.

Teorema

La successione $\{x_k\}$ converge ad una soluzione x^* di $F(x) = y$ tale che $\|x^* - x^\dagger\| \leq \rho$.

In questo caso si dimostra che $\lim_{k \rightarrow \infty} \|g_k\| = 0$ e di conseguenza il raggio di trust-region tende a zero.

Risultati per $\delta > 0$

Lemma

Sia $\bar{k} < k^*(\delta)$. Il metodo descritto genera una successione $\{x_k^\delta\}$ tale che per $\bar{k} \leq k < k^*(\delta)$

- la trust-region è attiva, cioè $\lambda_k > 0$;
- x_k^δ appartiene a $B_{2\rho}(x_{\bar{k}}^\delta)$ e a $B_\rho(x^\dagger)$;
- $\|x_{k+1}^\delta - x^\dagger\| < \|x_k^\delta - x^\dagger\|$;
- esiste una costante $\bar{\lambda} > 0$ tale che $\lambda_k \leq \bar{\lambda}$.

Teorema

Il principio di discrepanza viene soddisfatto dopo un numero finito $k^*(\delta)$ di iterazioni e la successione $\{x_{k^*(\delta)}^\delta\}$ converge ad una soluzione di $F(x) = y$ se δ tende a zero.

Il procedimento proposto è regolarizzante

Problemi test

- Quattro sistemi non lineari mal posti ottenuti dalla discretizzazione di equazioni integrali di Fredholm non lineari del primo tipo che modellizzano problemi di gravimetria e geofisica:

$$\int_0^1 k(t, s, x(s)) ds = y(t), \quad t \in [0, 1],$$

P1,P2, [Vogel, 1990], **P3,P4** [Kaltenbacher,2007];

- intervallo $[0, 1]$ discretizzato con $n=64$ punti equidistanti;
- λ_k calcolato risolvendo con il metodo di Newton

$$\psi(\lambda) = \frac{1}{\|p(\lambda)\|} - \frac{1}{\Delta_k} = 0.$$

Ogni iterazione del metodo di Newton richiede la fattorizzazione di Cholesky di $B_k + \lambda_k I$.

Implementazione trust-region regolarizzante

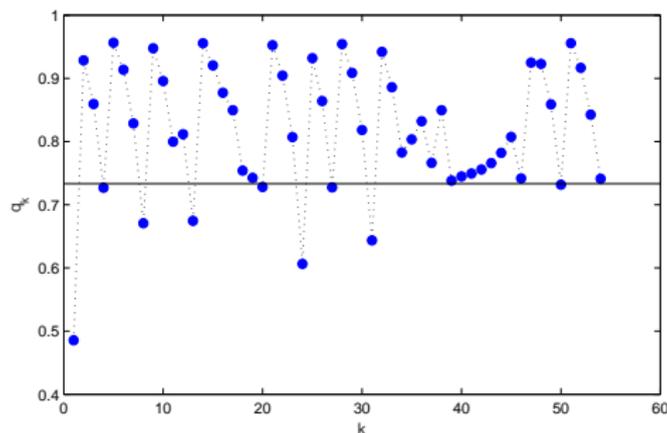
Aggiornamento raggio trust-region:

$$\Delta_k = \mu_k \|F(x_k^\delta) - y^\delta\|, \quad \mu_k = \begin{cases} \frac{1}{6}\mu_{k-1} & \text{se } q_{k-1} < q \\ 2\mu_{k-1} & \text{se } q_{k-1} > \nu q \\ \mu_{k-1} & \text{altrimenti} \end{cases}$$

con $q_k = \frac{\|F(x_k^\delta) - y^\delta + J(x_k^\delta)p_k\|}{\|F(x_k^\delta) - y^\delta\|}$, e $\nu = 1.1$.

- Δ_k è meno costoso da calcolare rispetto a $\frac{1-q}{\|B_k\|} \|g_k\|$ e mantiene la convergenza a zero quando $\delta = 0$.
- Nel suo aggiornamento si tiene conto del soddisfacimento della q-condition.

Proprietà regolarizzanti del metodo



- = Valori di $q_k = \frac{\|F(x_k^\delta) - y^\delta + J(x_k^\delta)p_k\|}{\|F(x_k^\delta) - y^\delta\|}$, linea continua: $q = 1.1/\tau$.

La q-condition è soddisfatta nella maggior parte delle iterazioni anche se non esplicitamente imposta.

Proprietà regolarizzanti del metodo

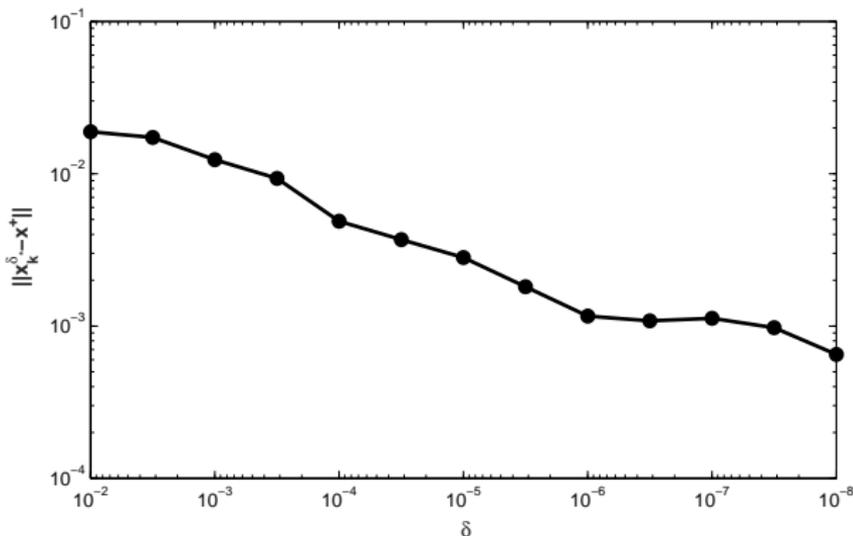


Grafico logaritmico dell'errore $\|x_{k^*}^\delta - x^\dagger\|$ in funzione del rumore δ .

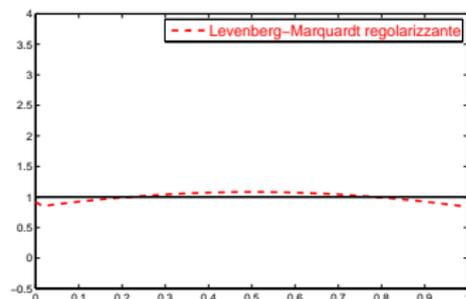
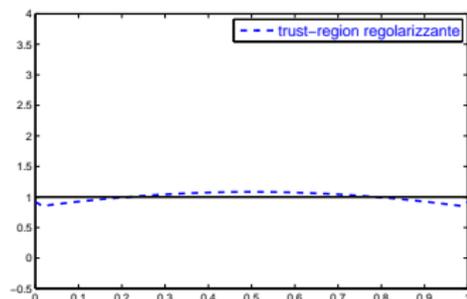
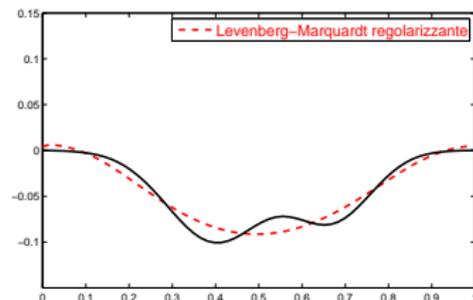
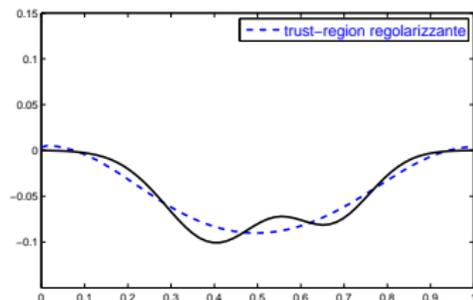
Confronto TR e LM regolarizzanti con rumore $\delta = 10^{-2}$.

Problem	x_0	TR regolarizzante			LM regolarizzante		
		it	nf	cf	it	nf	cf
P1	0 e	20	21	6	17	18	4
	-0.5 e	29	30	6	22	23	4
	-1 e	35	36	5	24	25	4
	-2 e	40	41	5	25	26	4
P2	0 e	30	31	5	*	*	*
	0.5 e	25	26	5	*	*	*
	1 e	29	30	5	22	23	5
	2 e	37	39	5	25	26	5
P3	$x_0(1.25)$	15	16	4	12	13	4
	$x_0(1.5)$	17	18	4	14	15	4
	$x_0(1.75)$	19	20	4	15	16	4
	$x_0(2)$	22	23	4	16	17	4
P4	$x_0(1, 1)$	17	18	5	10	11	4
	$x_0(0.5, 0)$	20	21	4	*	*	*
	$x_0(1.5, 1)$	22	23	4	15	16	4
	$x_0(1.5, 0)$	26	27	4	*	*	*

it=iterazioni,
nf=valutazioni
 funzione,
cf=numero medio
 fattorizzazioni
 cholesky.
 *=fallimento,
 raggiunto numero
 massimo di
 iterazioni o
 convergenza ad
 una soluzione del
 problema rumoroso

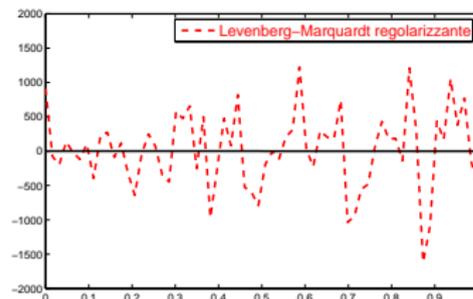
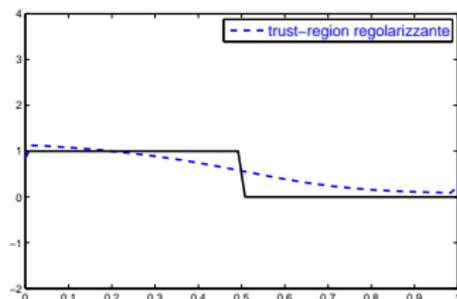
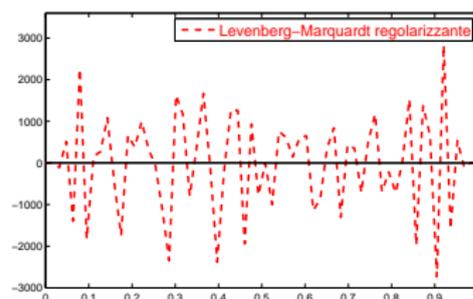
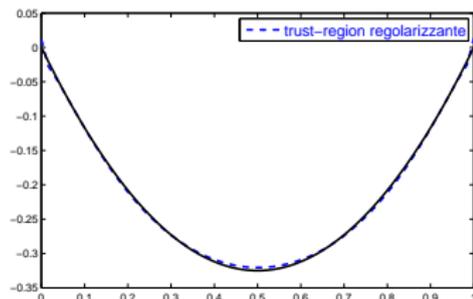
$e = (1, \dots, 1)^T$, **P3**: $(x_0(\alpha))_j = (-4\alpha + 4)s_j^2 + (4\alpha - 4)s_j + 1$, **P4**: $x_0(\beta, \chi) = \beta - \chi s_j$, s_j punti
 della griglia, $j = 1, \dots, n$.

Confronto TR e LM regolarizzanti



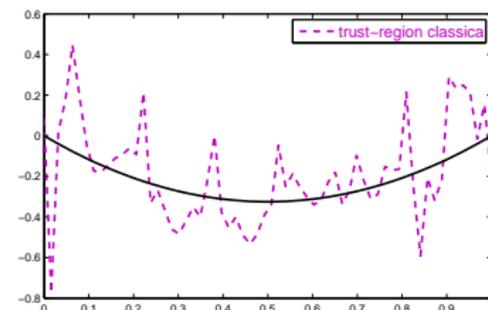
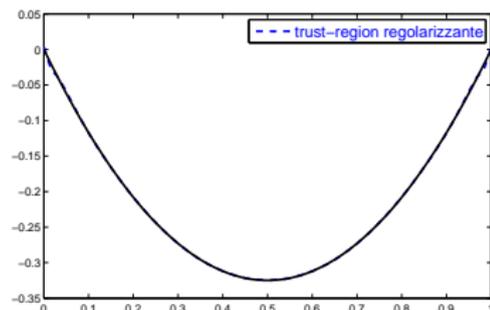
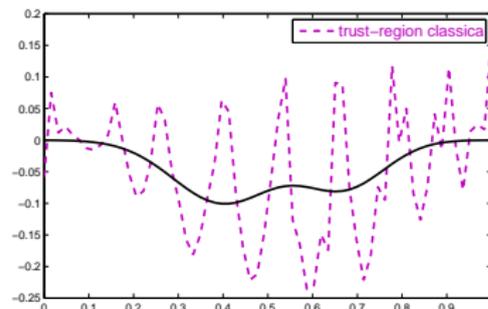
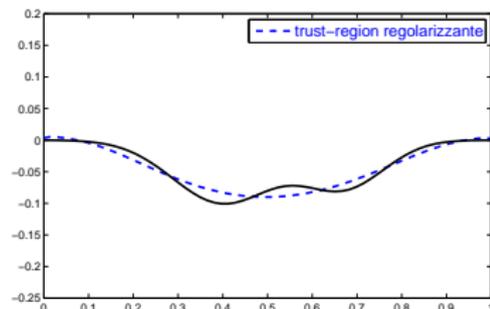
Sinistra: TR regolarizzante, **Destra:** LM regolarizzante, **Linea continua:** soluzione del problema originale.

Confronto TR e LM regolarizzanti



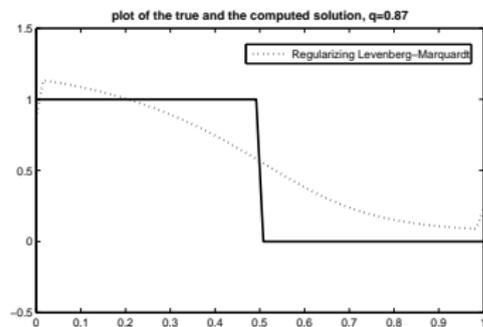
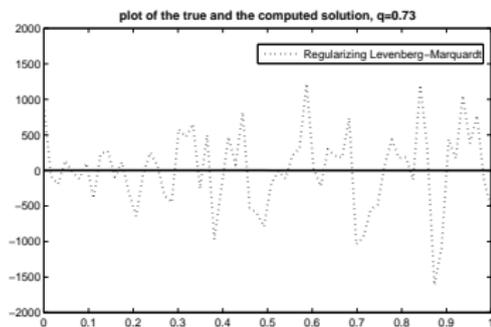
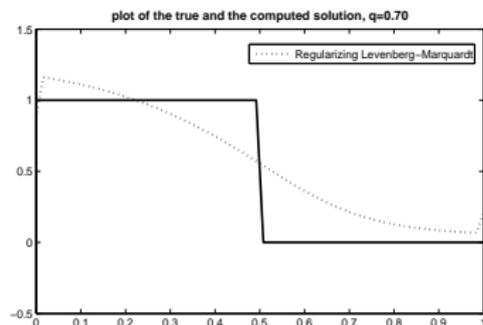
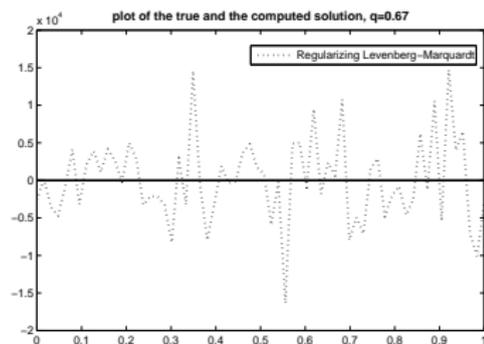
Sinistra: TR regolarizzante, **Destra:** LM regolarizzante, **Linea continua:** soluzione del problema originale.

Confronto trust-region regolarizzante e classica



Sinistra: TR regolarizzante, **Destra:** TR classica, **Linea continua:** soluzione del problema originale.

GRAZIE PER L'ATTENZIONE



Valori di $q = 0.67, 0.70, 0.73, 0.87$.

Confronto trust-region regolarizzante e classica

