

# Región de Confianza en Métodos Estructurados Secantes para Mínimos Cuadrados

Isidro A. Abelló Ugalde<sup>1</sup>

<sup>1</sup>Grupo de Nuevas Tecnologías en la Educación  
Centro de Estudios Para el Perfeccionamiento de la Educación Superior

Seminario de Dr. Pablo Barrera Sánchez  
Facultad de Ciencias, CU  
UNAM, 2009



# Esquema de la Presentación

- 1 Motivación
  - Formulación de los mínimos cuadrados no lineales
  - Métodos Clásicos de Solución
- 2 Métodos Estructurados y Región de Confianza
  - Principales Métodos Estructurados Cuasi-Newton
  - Subproblema de Región de Confianza



# Presentación

- 1 Motivación
  - Formulación de los mínimos cuadrados no lineales
  - Métodos Clásicos de Solución
- 2 Métodos Estructurados y Región de Confianza
  - Principales Métodos Estructurados Cuasi-Newton
  - Subproblema de Región de Confianza

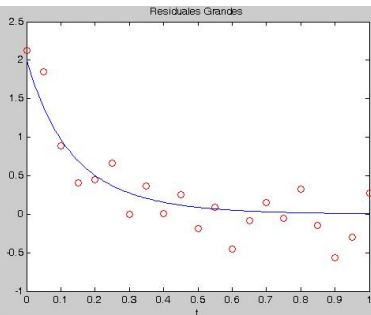
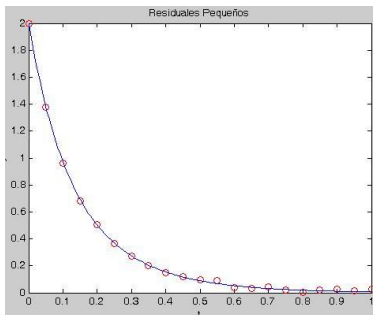


# Formulación

Mínimos Cuadrados No Lineales. El gradiente y el Hessiano.

$$\min_x f(x) = \frac{1}{2} \mathbf{r}(x)^t \mathbf{r}(x) = \frac{1}{2} \|\mathbf{r}(x)\|_2^2 = \frac{1}{2} \sum_{i=1}^m (r_i(x))^2,$$

donde:  $\mathbf{r} = \begin{bmatrix} r_1(x) \\ \vdots \\ r_m(x) \end{bmatrix}$ ,  $x \in \mathbb{R}^n$ ,  $m \geq n$



# Formulación

## Mínimos Cuadrados No Lineales. El gradiente y el Hessiano.

- **El gradiente:**  $\nabla f(x) = J(x)^t \mathbf{r}(x)$ ,

$$\text{donde } J = \begin{bmatrix} \frac{\partial r_1(x)}{\partial x_1} & \dots & \frac{\partial r_1(x)}{\partial x_n} \\ \frac{\partial r_2(x)}{\partial x_1} & \dots & \frac{\partial r_2(x)}{\partial x_n} \\ \vdots & \dots & \vdots \\ \vdots & \dots & \vdots \\ \frac{\partial r_m(x)}{\partial x_1} & \dots & \frac{\partial r_m(x)}{\partial x_n} \end{bmatrix}$$

- **El hessiano:**  $\nabla^2 f(x) = C(x) + S(x)$ ,

donde

- $C(x) \equiv J(x)^t J(x)$  Solo  $1^{ras}$  derivadas.
- $S(x) \equiv \sum_{i=1}^m r_i(x) \nabla^2 r_i(x)$  Solo  $\mathbf{r}$  y  $2^{das}$  derivadas.



# Formulación

## Mínimos Cuadrados No Lineales. El gradiente y el Hessiano.

- **El gradiente:**  $\nabla f(x) = J(x)^t \mathbf{r}(x)$ ,

$$\text{donde } J = \begin{bmatrix} \frac{\partial r_1(x)}{\partial x_1} & \dots & \frac{\partial r_1(x)}{\partial x_n} \\ \frac{\partial r_2(x)}{\partial x_1} & \dots & \frac{\partial r_2(x)}{\partial x_n} \\ \vdots & \dots & \vdots \\ \vdots & \dots & \vdots \\ \frac{\partial r_m(x)}{\partial x_1} & \dots & \frac{\partial r_m(x)}{\partial x_n} \end{bmatrix}$$

- **El hessiano:**  $\nabla^2 f(x) = C(x) + S(x)$ ,

donde

- $C(x) \equiv J(x)^t J(x)$  Solo  $1^{ras}$  derivadas.
- $S(x) \equiv \sum_{i=1}^m r_i(x) \nabla^2 r_i(x)$  Solo  $\mathbf{r}$  y  $2^{das}$  derivadas.



# Formulación

## Mínimos Cuadrados No Lineales. El gradiente y el Hessiano.

- **El gradiente:**  $\nabla f(x) = J(x)^t \mathbf{r}(x)$ ,

$$\text{donde } J = \begin{bmatrix} \frac{\partial r_1(x)}{\partial x_1} & \dots & \frac{\partial r_1(x)}{\partial x_n} \\ \frac{\partial r_2(x)}{\partial x_1} & \dots & \frac{\partial r_2(x)}{\partial x_n} \\ \vdots & \dots & \vdots \\ \vdots & \dots & \vdots \\ \frac{\partial r_m(x)}{\partial x_1} & \dots & \frac{\partial r_m(x)}{\partial x_n} \end{bmatrix}$$

- **El hessiano:**  $\nabla^2 f(x) = C(x) + S(x)$ ,

donde

- $C(x) \equiv J(x)^t J(x)$  Solo  $1^{ras}$  derivadas.
- $S(x) \equiv \sum_{i=1}^m r_i(x) \nabla^2 r_i(x)$  Solo  $\mathbf{r}$  y  $2^{das}$  derivadas.



# Ejemplo en Ajuste de Curvas

## Caso Particular

Ajustar  $A, \alpha, B, \beta$  del modelo  $y = Ae^{-\alpha t} + Be^{-\beta t} + \epsilon(0; \sigma^2)$

de una tabla de datos  $(t_i; y_i)$ , con  $1 \leq i \leq m$

$$x = [A, \alpha, B, \beta]^t \quad n = 4$$

$$r(x) = \begin{bmatrix} y_1 - x_1 e^{-x_2 t_1} - x_3 e^{-x_4 t_1} \\ \vdots \\ y_m - x_1 e^{-x_2 t_m} - x_3 e^{-x_4 t_m} \end{bmatrix}_{(m \times 1)}, \quad f(x) = \frac{1}{2} r^t r$$

$$J(x) = \begin{bmatrix} -e^{-x_2 t_1} & x_1 t_1 e^{-x_2 t_1} & -e^{-x_4 t_1} & x_3 t_1 e^{-x_4 t_1} \\ \vdots & \vdots & \vdots & \vdots \\ -e^{-x_2 t_m} & x_1 t_m e^{-x_2 t_m} & -e^{-x_4 t_m} & x_3 t_m e^{-x_4 t_m} \end{bmatrix}_{(m \times 4)}, \quad \nabla f = J^t r$$

$$\nabla^2 r_i(x) = \begin{bmatrix} 0 & t_i e^{-x_2 t_i} & 0 & 0 \\ -t_i e^{-x_2 t_i} & -x_1 t_i^2 e^{-x_2 t_i} & 0 & 0 \\ 0 & 0 & 0 & t_i e^{-x_4 t_i} \\ 0 & 0 & -t_i e^{-x_4 t_i} & -x_3 t_i^2 e^{-x_4 t_i} \end{bmatrix}_{(4 \times 4)}, \quad \nabla^2 f = J^t J + \sum_{i=1}^m r_i \nabla^2 r_i$$





# Presentación

- 1 Motivación
  - Formulación de los mínimos cuadrados no lineales
  - Métodos Clásicos de Solución
- 2 Métodos Estructurados y Región de Confianza
  - Principales Métodos Estructurados Cuasi-Newton
  - Subproblema de Región de Confianza



# Métodos Clásicos

## Gauss-Newton y Levenberg-Marquardt.

Si  $S(x) \equiv \sum_{i=1}^m r_i(x) \nabla^2 r_i(x) \approx 0$  then  $\nabla^2 f \approx J^t J$

- Gauss-Newton (modificado con *line search*)

- $x_+ = x_c + \alpha s_c$
- $s_c = \min_p \{ \|Jp + r\|_2^2 \} = -(J^t J)^{-1} J^t r$
- $\alpha$  es el tamaño de paso

- Levenberg-Marquardt (región de confianza)

- $x_+ = x_c + s_c$
- $s_c = \min_p \{ \|Jp + r\|_2^2 : \|Dp\| \leq \Delta \}$

$$(J^t J + \lambda D^t D) s_c = -J^t r$$



# Métodos Clásicos

## Gauss-Newton y Levenberg-Marquardt.

Si  $S(x) \equiv \sum_{i=1}^m r_i(x) \nabla^2 r_i(x) \approx 0$  then  $\nabla^2 f \approx J^t J$

- **Gauss-Newton (modificado con *line search*)**

- $x_+ = x_c + \alpha s_c$
- $s_c = \min_p \left\{ \|Jp + \mathbf{r}\|_2^2 \right\} = - (J^t J)^{-1} J^t \mathbf{r}$
- $\alpha$  es el tamaño de paso

- **Levenberg-Marquardt (región de confianza)**

- $x_+ = x_c + s_c$
- $s_c = \min_p \left\{ \|Jp + \mathbf{r}\|_2^2 : \|Dp\| \leq \Delta \right\}$

$$(J^t J + \lambda D^t D) s_c = -J^t \mathbf{r}$$



# Métodos Clásicos

## Gauss-Newton y Levenberg-Marquardt.

Si  $S(x) \equiv \sum_{i=1}^m r_i(x) \nabla^2 r_i(x) \approx 0$  then  $\nabla^2 f \approx J^t J$

- **Gauss-Newton (modificado con *line search*)**

- $x_+ = x_c + \alpha s_c$
- $s_c = \min_p \left\{ \|Jp + r\|_2^2 \right\} = - (J^t J)^{-1} J^t r$
- $\alpha$  es el tamaño de paso

- **Levenberg-Marquardt (región de confianza)**

- $x_+ = x_c + s_c$
- $s_c = \min_p \left\{ \|Jp + r\|_2^2 : \|Dp\| \leq \Delta \right\}$

$$(J^t J + \lambda D^t D) s_c = -J^t r$$



# Métodos Clásicos

## Gauss-Newton y Levenberg-Marquardt.

Si  $S(x) \equiv \sum_{i=1}^m r_i(x) \nabla^2 r_i(x) \approx 0$  then  $\nabla^2 f \approx J^t J$

- **Gauss-Newton (modificado con *line search*)**

- $x_+ = x_c + \alpha s_c$
- $s_c = \min_p \left\{ \|Jp + \mathbf{r}\|_2^2 \right\} = - (J^t J)^{-1} J^t \mathbf{r}$
- $\alpha$  es el tamaño de paso

- **Levenberg-Marquardt (región de confianza)**

- $x_+ = x_c + s_c$
- $s_c = \min_p \left\{ \|Jp + \mathbf{r}\|_2^2 : \|Dp\| \leq \Delta \right\}$

$$(J^t J + \lambda D^t D) s_c = -J^t \mathbf{r}$$



# Métodos Clásicos

## Gauss-Newton y Levenberg-Marquardt.

Si  $S(x) \equiv \sum_{i=1}^m r_i(x) \nabla^2 r_i(x) \approx 0$  then  $\nabla^2 f \approx J^t J$

- **Gauss-Newton (modificado con *line search*)**

- $x_+ = x_c + \alpha s_c$
- $s_c = \min_p \left\{ \|Jp + \mathbf{r}\|_2^2 \right\} = - (J^t J)^{-1} J^t \mathbf{r}$
- $\alpha$  es el tamaño de paso

- **Levenberg-Marquardt (región de confianza)**

- $x_+ = x_c + s_c$
- $s_c = \min_p \left\{ \|Jp + \mathbf{r}\|_2^2 : \|Dp\| \leq \Delta \right\}$

$$(J^t J + \lambda D^t D) s_c = -J^t \mathbf{r}$$



# Métodos Clásicos

## Gauss-Newton y Levenberg-Marquardt.

Si  $S(x) \equiv \sum_{i=1}^m r_i(x) \nabla^2 r_i(x) \approx 0$  then  $\nabla^2 f \approx J^t J$

- **Gauss-Newton (modificado con *line search*)**

- $x_+ = x_c + \alpha s_c$
- $s_c = \min_p \left\{ \|Jp + \mathbf{r}\|_2^2 \right\} = - (J^t J)^{-1} J^t \mathbf{r}$
- $\alpha$  es el tamaño de paso

- **Levenberg-Marquardt (región de confianza)**

- $x_+ = x_c + s_c$
- $s_c = \min_p \left\{ \|Jp + \mathbf{r}\|_2^2 : \|Dp\| \leq \Delta \right\}$

$$(J^t J + \lambda D^t D) s_c = -J^t \mathbf{r}$$



# Métodos Clásicos

## Gauss-Newton y Levenberg-Marquardt.

Si  $S(x) \equiv \sum_{i=1}^m r_i(x) \nabla^2 r_i(x) \approx 0$  then  $\nabla^2 f \approx J^t J$

- **Gauss-Newton (modificado con *line search*)**

- $x_+ = x_c + \alpha s_c$
- $s_c = \min_p \left\{ \|Jp + \mathbf{r}\|_2^2 \right\} = - (J^t J)^{-1} J^t \mathbf{r}$
- $\alpha$  es el tamaño de paso

- **Levenberg-Marquardt (región de confianza)**

- $x_+ = x_c + s_c$
- $s_c = \min_p \left\{ \|Jp + \mathbf{r}\|_2^2 : \|Dp\| \leq \Delta \right\}$

$$\begin{array}{c} \Updownarrow \\ (J^t J + \lambda D^t D) s_c = -J^t \mathbf{r} \end{array}$$





# Métodos Clásicos

## Propiedades y Deficiencias

- **Propiedades:**

- Convergencia q-lineal local
- Buenas estimaciones del hessiano real y rápidas tasas de convergencia si
  - 1 la función es cercana a ser lineal (los términos de segundo orden son pequeños).
  - 2 el vector residual es cercano a cero.
- Convergencia q-cuadrática si tenemos un caso cero-residual.

- **Limitaciones:**

- pobre desempeño cuando los residuales son grandes en la solución, o la función es muy no lineal. Es estos casos la tasa de convergencia cae de q-cuadrática a q-lineal.
- No hay garantías de convergencia global.



# Métodos Clásicos

## Propiedades y Deficiencias

- **Propiedades:**
  - **Convergencia q-lineal local**
  - Buenas estimaciones del hessiano real y rápidas tasas de convergencia si
    - 1 la función es cercana a ser lineal (los términos de segundo orden son pequeños).
    - 2 el vector residual es cercano a cero.
  - Convergencia q-cuadrática si tenemos un caso cero-residual.
- **Limitaciones:**
  - pobre desempeño cuando los residuales son grandes en la solución, o la función es muy no lineal. Es estos casos la tasa de convergencia cae de q-cuadrática a q-lineal.
  - No hay garantías de convergencia global.



# Métodos Clásicos

## Propiedades y Deficiencias

- Propiedades:
  - Convergencia q-lineal local
  - Buenas estimaciones del hessiano real y rápidas tasas de convergencia si
    - 1 la función es cercana a ser lineal (los términos de segundo orden son pequeños).
    - 2 el vector residual es cercano a cero.
  - Convergencia q-cuadrática si tenemos un caso cero-residual.
- Limitaciones:
  - pobre desempeño cuando los residuales son grandes en la solución, o la función es muy no lineal. Es estos casos la tasa de convergencia cae de q-cuadrática a q-lineal.
  - No hay garantías de convergencia global.



# Métodos Clásicos

## Propiedades y Deficiencias

- Propiedades:
  - Convergencia q-lineal local
  - Buenas estimaciones del hessiano real y rápidas tasas de convergencia si
    - 1 la función es cercana a ser lineal (los términos de segundo orden son pequeños).
    - 2 el vector residual es cercano a cero.
  - Convergencia q-cuadrática si tenemos un caso cero-residual.
- Limitaciones:
  - pobre desempeño cuando los residuales son grandes en la solución, o la función es muy no lineal. Es estos casos la tasa de convergencia cae de q-cuadrática a q-lineal.
  - No hay garantías de convergencia global.



# Métodos Clásicos

## Propiedades y Deficiencias

- Propiedades:
  - Convergencia q-lineal local
  - Buenas estimaciones del hessiano real y rápidas tasas de convergencia si
    - 1 la función es cercana a ser lineal (los términos de segundo orden son pequeños).
    - 2 el vector residual es cercano a cero.
  - Convergencia q-cuadrática si tenemos un caso cero-residual.
- Limitaciones:
  - pobre desempeño cuando los residuales son grandes en la solución, o la función es muy no lineal. Es estos casos la tasa de convergencia cae de q-cuadrática a q-lineal.
  - No hay garantías de convergencia global.



# Métodos Clásicos

## Propiedades y Deficiencias

- Propiedades:
  - Convergencia q-lineal local
  - Buenas estimaciones del hessiano real y rápidas tasas de convergencia si
    - 1 la función es cercana a ser lineal (los términos de segundo orden son pequeños).
    - 2 el vector residual es cercano a cero.
  - Convergencia q-cuadrática si tenemos un caso cero-residual.
- Limitaciones:
  - pobre desempeño cuando los residuales son grandes en la solución, o la función es muy no lineal. Es estos casos la tasa de convergencia cae de q-cuadrática a q-lineal.
  - No hay garantías de convergencia global.



# Presentación

- 1 Motivación
  - Formulación de los mínimos cuadrados no lineales
  - Métodos Clásicos de Solución
- 2 Métodos Estructurados y Región de Confianza
  - Principales Métodos Estructurados Cuasi-Newton
  - Subproblema de Región de Confianza



# Principales Métodos Estructurados Cuasi-Newton

## Método Dennis, Gay and Welsh (DWG)

$$H_k = J^t J + B_k \approx \nabla^2 f(x_k)$$

$$\text{donde } B_k \approx \sum_{j=1}^m r_j(x_k) \nabla^2 r_j(x_k)$$

Consideremos la ecuación secante de cada matriz hessiana  $\nabla^2 r_j(x)$

$$(B_j)_{k+1} (x_{k+1} - x_k) = \nabla r_j(x_{k+1}) - \nabla r_j(x_k)$$

$$\begin{aligned} \mathbf{B}_{k+1} (x_{k+1} - x_k) &= \sum_{j=1}^m r_j(x_{k+1}) (B_j)_{k+1} (x_{k+1} - x_k) \\ &= \sum_{j=1}^m r_j(x_{k+1}) (\nabla r_j(x_{k+1}) - \nabla r_j(x_k)) \\ &= (J_{k+1} - J_k)^t \mathbf{r}_{k+1} \end{aligned}$$

La nueva ecuación secante:  $\boxed{B_{k+1} s_k = y^\#}$  donde  $y^\# = (J_{k+1} - J_k)^t \mathbf{r}_{k+1}$

$\mathbf{B}_{k+1}$  se encuentra por algún miembro de la familia convexa de Broyden.  
Ej. BFGS, DFP, PSB, etc.





# Principales Métodos Estructurados Cuasi-Newton

## Método Dennis, Gay and Welsh (DWG)

$$H_k = J^t J + B_k \approx \nabla^2 f(x_k)$$

$$\text{donde } B_k \approx \sum_{j=1}^m r_j(x_k) \nabla^2 r_j(x_k)$$

Consideremos la ecuación secante de cada matriz hessiana  $\nabla^2 r_j(x)$

$$(B_j)_{k+1} (x_{k+1} - x_k) = \nabla r_j(x_{k+1}) - \nabla r_j(x_k)$$

$$\begin{aligned} B_{k+1} (x_{k+1} - x_k) &= \sum_{j=1}^m r_j(x_{k+1}) (B_j)_{k+1} (x_{k+1} - x_k) \\ &= \sum_{j=1}^m r_j(x_{k+1}) (\nabla r_j(x_{k+1}) - \nabla r_j(x_k)) \\ &= (J_{k+1} - J_k)^t r_{k+1} \end{aligned}$$

La nueva ecuación secante:  $B_{k+1} s_k = y^\#$  donde  $y^\# = (J_{k+1} - J_k)^t r_{k+1}$

$B_{k+1}$  se encuentra por algún miembro de la familia convexa de Broyden.  
 Ej. BFGS, DFP, PSB, etc.



# Principales Métodos Estructurados Cuasi-Newton

## Método Dennis, Gay and Welsh (DWG)

$$H_k = J^t J + B_k \approx \nabla^2 f(x_k)$$

$$\text{donde } B_k \approx \sum_{j=1}^m r_j(x_k) \nabla^2 r_j(x_k)$$

Consideremos la ecuación secante de cada matriz hessiana  $\nabla^2 r_j(x)$

$$(B_j)_{k+1} (x_{k+1} - x_k) = \nabla r_j(x_{k+1}) - \nabla r_j(x_k)$$

$$\begin{aligned} \mathbf{B}_{k+1} (x_{k+1} - x_k) &= \sum_{j=1}^m r_j(x_{k+1}) (B_j)_{k+1} (x_{k+1} - x_k) \\ &= \sum_{j=1}^m r_j(x_{k+1}) (\nabla r_j(x_{k+1}) - \nabla r_j(x_k)) \\ &= (\mathbf{J}_{k+1} - \mathbf{J}_k)^t \mathbf{r}_{k+1} \end{aligned}$$

La nueva ecuación secante:  $\boxed{\mathbf{B}_{k+1} s_k = y^\#}$  donde  $y^\# = (\mathbf{J}_{k+1} - \mathbf{J}_k)^t \mathbf{r}_{k+1}$

$\mathbf{B}_{k+1}$  se encuentra por algún miembro de la familia convexa de Broyden.  
 Ej. BFGS, DFP, PSB, etc.



# Método de Dennis, Gay, Welsh

Propiedades y Limitaciones. Alternativa de Fletcher & Xu

## ● Propiedades:

- convergencia q-superlineal.
- razonable estimación de la matriz Hessiana real.
- No garantiza que las actualizaciones  $B_k$  sean definidas positivas (Conduce a región de confianza).

## ● Limitación:

- $B_k$ , no garantiza convergencia a la matriz nula en el caso cero residual. Se pierden las rápidas tasas de convergencia de GN y LM para este caso.

### Solución de Fletcher & Xu: Métodos Híbridos

Controlar el residual y utilizar Gauss-Newton en residuales pequeños y BFGS estructurado en residuales grandes.

**Punto Crítico:** La política de cambio entre dos diferentes algoritmos es crucial para lograr tasas de convergencia de q-superlineal a q-cuadrática.



# Método de Dennis, Gay, Welsh

Propiedades y Limitaciones. Alternativa de Fletcher & Xu

## ● Propiedades:

- convergencia q-superlineal.
- razonable estimación de la matriz Hessiana real.
- No garantiza que las actualizaciones  $B_k$  sean definidas positivas (Conduce a región de confianza).

## ● Limitación:

- $B_k$ , no garantiza convergencia a la matriz nula en el caso cero residual. Se pierden las rápidas tasas de convergencia de GN y LM para este caso.

### Solución de Fletcher & Xu: Métodos Híbridos

Controlar el residual y utilizar Gauss-Newton en residuales pequeños y BFGS estructurado en residuales grandes.

**Punto Crítico:** La política de cambio entre dos diferentes algoritmos es crucial para lograr tasas de convergencia de q-superlineal a q-cuadrática.



# Método de Dennis, Gay, Welsh

Propiedades y Limitaciones. Alternativa de Fletcher & Xu

## ● Propiedades:

- convergencia q-superlineal.
- razonable estimación de la matriz Hessiana real.
- No garantiza que las actualizaciones  $B_k$  sean definidas positivas (Conduce a región de confianza).

## ● Limitación:

- $B_k$ , no garantiza convergencia a la matriz nula en el caso cero residual. Se pierden las rápidas tasas de convergencia de GN y LM para este caso.

### Solución de Fletcher & Xu: Métodos Híbridos

Controlar el residual y utilizar Gauss-Newton en residuales pequeños y BFGS estructurado en residuales grandes.

**Punto Crítico:** La política de cambio entre dos diferentes algoritmos es crucial para lograr tasas de convergencia de q-superlineal a q-cuadrática.



# Método de Dennis, Gay, Welsh

Propiedades y Limitaciones. Alternativa de Fletcher & Xu

## ● Propiedades:

- convergencia q-superlineal.
- razonable estimación de la matriz Hessiana real.
- No garantiza que las actualizaciones  $B_k$  sean definidas positivas (Conduce a región de confianza).

## ● Limitación:

- $B_k$ , no garantiza convergencia a la matriz nula en el caso cero residual. Se pierden las rápidas tasas de convergencia de GN y LM para este caso.

### Solución de Fletcher & Xu: Métodos Híbridos

Controlar el residual y utilizar Gauss-Newton en residuales pequeños y BFGS estructurado en residuales grandes.

**Punto Crítico:** La política de cambio entre dos diferentes algoritmos es crucial para lograr tasas de convergencia de q-superlineal a q-cuadrática.



# Método de Dennis, Gay, Welsh

Propiedades y Limitaciones. Alternativa de Fletcher & Xu

## ● Propiedades:

- convergencia q-superlineal.
- razonable estimación de la matriz Hessiana real.
- No garantiza que las actualizaciones  $B_k$  sean definidas positivas (Conduce a región de confianza).

## ● Limitación:

- $B_k$ , no garantiza convergencia a la matriz nula en el caso cero residual. Se pierden las rápidas tasas de convergencia de GN y LM para este caso.

### Solución de Fletcher & Xu: Métodos Híbridos

Controlar el residual y utilizar Gauss-Newton en residuales pequeños y BFGS estructurado en residuales grandes.

**Punto Crítico:** La política de cambio entre dos diferentes algoritmos es crucial para lograr tasas de convergencia de q-superlineal a q-cuadrática.



# Método de Dennis, Gay, Welsh

Propiedades y Limitaciones. Alternativa de Fletcher & Xu

## ● Propiedades:

- convergencia q-superlineal.
- razonable estimación de la matriz Hessiana real.
- No garantiza que las actualizaciones  $B_k$  sean definidas positivas (Conduce a región de confianza).

## ● Limitación:

- $B_k$ , no garantiza convergencia a la matriz nula en el caso cero residual. Se pierden las rápidas tasas de convergencia de GN y LM para este caso.

### **Solución de Fletcher & Xu: Métodos Híbridos**

Controlar el residual y utilizar Gauss-Newton en residuales pequeños y BFGS estructurado en residuales grandes.

**Punto Crítico:** La política de cambio entre dos diferentes algoritmos es crucial para lograr tasas de convergencia de q-superlineal a q-cuadrática.





# Método de Huschens: Totalmente Estructurado

## Idea del método. Propiedades

Reescribir  $\nabla^2 f(x)$

$$\nabla^2 f(x) = J(x)^t J(x) + \|r(x)\| \sum_{i=1}^m \frac{r_i(x)}{\|r(x)\|} \nabla^2 r_i(x)$$

$$H_+ = J_+^t J_+ + \|r\| B_+ \quad \text{ahora} \quad B_+ \approx \sum_{j=1}^m \frac{r_j(x_+)}{\|r\|} \nabla^2 r_j(x_+)$$

$$(B_j)_+ \approx \nabla^2 r_j(x_+) \text{ vía ecuación secante } (B_j)_+ s_c = \nabla r_j(x_+) - \nabla r_j(x_c)$$

$$\begin{aligned} B_+ s_c &= \sum_{j=1}^m \frac{r_j(x_+)}{\|r\|} (B_j)_+ s_c \\ &= \sum_{j=1}^m \frac{r_j(x_+)}{\|r\|} (\nabla r_j(x_+) - \nabla r_j(x_c)) \\ &= (J_+ - J_c)^t \frac{r_+}{\|r\|} = \tilde{y}^\# \end{aligned}$$

**Propiedades:**

- autoescalamiento sin política de *switching*  $(\frac{r_i(x)}{\|r(x)\|})$ .
- convergencia q-cuadrática para el caso cero-residual.
- convergencia q-superlineal para el caso residual mayor que cero.
- Buena estimación de la matriz Hessiana en el óptimo.



# Método de Huschens: Totalmente Estructurado

## Idea del método. Propiedades

Reescribir  $\nabla^2 f(x)$

$$\nabla^2 f(x) = J(x)^t J(x) + \|\mathbf{r}(x)\| \sum_{i=1}^m \frac{r_i(x)}{\|\mathbf{r}(x)\|} \nabla^2 r_i(x)$$

$$H_+ = J_+^t J_+ + \|\mathbf{r}\| \mathbf{B}_+ \quad \text{ahora} \quad \mathbf{B}_+ \approx \sum_{j=1}^m \frac{r_j(x_+)}{\|\mathbf{r}\|} \nabla^2 r_j(x_+)$$

$$(B_j)_+ \approx \nabla^2 r_j(x_+) \quad \text{vía ecuación secante} \quad (B_j)_+ \mathbf{s}_c = \nabla r_j(x_+) - \nabla r_j(x_c)$$

$$\begin{aligned} \mathbf{B}_+ \mathbf{s}_c &= \sum_{j=1}^m \frac{r_j(x_+)}{\|\mathbf{r}\|} (B_j)_+ \mathbf{s}_c \\ &= \sum_{j=1}^m \frac{r_j(x_+)}{\|\mathbf{r}\|} (\nabla r_j(x_+) - \nabla r_j(x_c)) \\ &= (J_+ - J_c)^t \frac{\mathbf{r}_+}{\|\mathbf{r}\|} = \tilde{\mathbf{y}}^\# \end{aligned}$$

Propiedades:

- autoescalamiento sin política de *switching*  $(\frac{r_i(x)}{\|\mathbf{r}(x)\|})$ .
- convergencia q-cuadrática para el caso cero-residual.
- convergencia q-superlineal para el caso residual mayor que cero.
- Buena estimación de la matriz Hessiana en el óptimo.



# Método de Huschens: Totalmente Estructurado

## Idea del método. Propiedades

Reescribir  $\nabla^2 f(x)$

$$\nabla^2 f(x) = J(x)^t J(x) + \|\mathbf{r}(x)\| \sum_{i=1}^m \frac{r_i(x)}{\|\mathbf{r}(x)\|} \nabla^2 r_i(x)$$

$$H_+ = J_+^t J_+ + \|\mathbf{r}\| \mathbf{B}_+ \quad \text{ahora} \quad \mathbf{B}_+ \approx \sum_{j=1}^m \frac{r_j(x_+)}{\|\mathbf{r}\|} \nabla^2 r_j(x_+)$$

$$(B_j)_+ \approx \nabla^2 r_j(x_+) \text{ vía ecuación secante } (B_j)_+ \mathbf{s}_c = \nabla r_j(x_+) - \nabla r_j(x_c)$$

$$\begin{aligned} \mathbf{B}_+ \mathbf{s}_c &= \sum_{j=1}^m \frac{r_j(x_+)}{\|\mathbf{r}\|} (B_j)_+ \mathbf{s}_c \\ &= \sum_{j=1}^m \frac{r_j(x_+)}{\|\mathbf{r}\|} (\nabla r_j(x_+) - \nabla r_j(x_c)) \\ &= (J_+ - J_c)^t \frac{\mathbf{r}_+}{\|\mathbf{r}\|} = \tilde{\mathbf{y}}^\# \end{aligned}$$

### Propiedades:

- autoescalamiento sin política de *switching*  $(\frac{r_i(x)}{\|\mathbf{r}(x)\|})$ .
- convergencia q-cuadrática para el caso cero-residual.
- convergencia q-superlineal para el caso residual mayor que cero.
- Buena estimación de la matriz Hessiana en el óptimo.



# Otros métodos totalmente estructurados

## Método de Zhang. Algoritmo

Zhang [2003] desarrolló una nueva ecuación secante  $\tilde{y}_Z^\#$

$$\begin{aligned} \tilde{y}_Z^\# &= (J_+ - J_c)^t \mathbf{r}_+ + \left(\frac{3\mathbf{r}_+}{s^t s}\right)^t [(J_+ + J_c) \mathbf{s} - 2(\mathbf{r}_+ - \mathbf{r}_c)] \mathbf{s} \\ &= \mathbf{y}^\# + \left(\frac{3\mathbf{r}_+}{s^t s}\right)^t [(J_+ + J_c) \mathbf{s} - 2(\mathbf{r}_+ - \mathbf{r}_c)] \mathbf{s} \end{aligned}$$

**Algoritmo GSSM**( $\Theta, y$ )(*Global*.)

Dado un  $x = x_0$  inicial y una matriz simétrica inicial  $\mathbf{B} = \mathbf{B}_0$

- 1 Sea  $H = J^t J + \Theta \mathbf{B}$  y resuelva el subproblema de región de confianza  $\min \{ \psi(s) = \nabla f(x)^t s + \frac{1}{2} s^t H s : \|s\| \leq \Delta \}$  para un  $\Delta$  aceptable
- 2  $x \leftarrow x + s$
- 3 (criterio de parada) Si  $\|\nabla f(x)\| < tol_1$  or  $\|s\| < tol_2$  parar.
- 4 (Actualización BFGS de  $\mathbf{B}$ )  $\mathbf{B} \leftarrow \mathbf{B} + \frac{\mathbf{B} s s^t \mathbf{B}}{s^t \mathbf{B} s} - \frac{y y^t}{y^t s}$  e ir al paso 1.

$$\text{GSSM}(\Theta; y) = \begin{cases} \Theta = 1 & y = \mathbf{y}^\# = (J_+ - J_c)^t \mathbf{r}_+ \\ \Theta = \|\mathbf{r}\| & y = \tilde{\mathbf{y}}^\# = \frac{\mathbf{y}^\#}{\Theta} \\ \Theta = \|\mathbf{r}\| & y = \tilde{\tilde{\mathbf{y}}}^\# \end{cases}$$

Dennis, Gay & Welsh  
 Huschens  
 Zhang



# Otros métodos totalmente estructurados

## Método de Zhang. Algoritmo

Zhang [2003] desarrolló una nueva ecuación secante  $\tilde{y}_Z^\#$

$$\begin{aligned}\tilde{y}_Z^\# &= (J_+ - J_c)^t \mathbf{r}_+ + \left(\frac{3\mathbf{r}_+}{s^t s}\right)^t [(J_+ + J_c) \mathbf{s} - 2(\mathbf{r}_+ - \mathbf{r}_c)] \mathbf{s} \\ &= \mathbf{y}^\# + \left(\frac{3\mathbf{r}_+}{s^t s}\right)^t [(J_+ + J_c) \mathbf{s} - 2(\mathbf{r}_+ - \mathbf{r}_c)] \mathbf{s}\end{aligned}$$

### Algoritmo GSSM( $\Theta, y$ )(Global.)

Dado un  $x = x_0$  inicial y una matriz simétrica inicial  $\mathbf{B} = \mathbf{B}_0$

- 1 Sea  $H = J^t J + \Theta \mathbf{B}$  y resuelva el subproblema de región de confianza  $\min \{ \psi(s) = \nabla f(x)^t s + \frac{1}{2} s^t H s : \|s\| \leq \Delta \}$  para un  $\Delta$  aceptable
- 2  $x \leftarrow x + s$
- 3 (criterio de parada) Si  $\|\nabla f(x)\| < tol_1$  or  $\|s\| < tol_2$  parar.
- 4 (Actualización BFGS de  $\mathbf{B}$ )  $\mathbf{B} \leftarrow \mathbf{B} + \frac{\mathbf{B} s s^t \mathbf{B}}{s^t \mathbf{B} s} - \frac{y y^t}{y^t s}$  e ir al paso 1.

$$\text{GSSM}(\Theta; y) = \begin{cases} \Theta = 1 & y = y^\# = (J_+ - J_c)^t \mathbf{r}_+ \\ \Theta = \|\mathbf{r}\| & y = \tilde{y}^\# = \frac{y^\#}{\Theta} \\ \Theta = \|\mathbf{r}\| & y = \tilde{y}_Z^\# \end{cases}$$

Dennis, Gay & Welsh  
 Huschens  
 Zhang



# Otros métodos totalmente estructurados

## Método de Zhang. Algoritmo

Zhang [2003] desarrolló una nueva ecuación secante  $\tilde{y}_Z^\#$

$$\begin{aligned}\tilde{y}_Z^\# &= (J_+ - J_c)^t \mathbf{r}_+ + \left(\frac{3\mathbf{r}_+}{s^t s}\right)^t [(J_+ + J_c) \mathbf{s} - 2(\mathbf{r}_+ - \mathbf{r}_c)] \mathbf{s} \\ &= \mathbf{y}^\# + \left(\frac{3\mathbf{r}_+}{s^t s}\right)^t [(J_+ + J_c) \mathbf{s} - 2(\mathbf{r}_+ - \mathbf{r}_c)] \mathbf{s}\end{aligned}$$

**Algoritmo GSSM**( $\Theta, y$ )(*Global.*)

Dado un  $x = x_0$  inicial y una matriz simétrica inicial  $\mathbf{B} = \mathbf{B}_0$

- 1 Sea  $H = J^t J + \Theta \mathbf{B}$  y resuelva el subproblema de región de confianza  $\min \{ \psi(\mathbf{s}) = \nabla f(x)^t \mathbf{s} + \frac{1}{2} \mathbf{s}^t H \mathbf{s} : \|\mathbf{s}\| \leq \Delta \}$  para un  $\Delta$  aceptable
- 2  $x \leftarrow x + \mathbf{s}$
- 3 (criterio de parada) Si  $\|\nabla f(x)\| < tol_1$  or  $\|\mathbf{s}\| < tol_2$  parar.
- 4 (Actualización BFGS de  $\mathbf{B}$ )  $\mathbf{B} \leftarrow \mathbf{B} + \frac{\mathbf{B} \mathbf{s} \mathbf{s}^t \mathbf{B}}{s^t \mathbf{B} \mathbf{s}} - \frac{y y^t}{y^t \mathbf{s}}$  e ir al paso 1.

$$\text{GSSM}(\Theta; y) = \begin{cases} \Theta = 1 & y = \mathbf{y}^\# = (J_+ - J_c)^t \mathbf{r}_+ \\ \Theta = \|\mathbf{r}\| & y = \tilde{\mathbf{y}}^\# = \frac{y^\#}{\Theta} \\ \Theta = \|\mathbf{r}\| & y = \tilde{\tilde{\mathbf{y}}}_Z^\# \end{cases}$$

Dennis, Gay & Welsh  
 Huschens  
 Zhang



# Presentación

- 1 Motivación
  - Formulación de los mínimos cuadrados no lineales
  - Métodos Clásicos de Solución
- 2 Métodos Estructurados y Región de Confianza
  - Principales Métodos Estructurados Cuasi-Newton
  - Subproblema de Región de Confianza



# Clase de Algoritmos de Región de Confianza

## Cuadrática restringida a una bola (elipse).

$$\min_w \{g^t w + \frac{1}{2} w^t B w : \|w\| \leq \Delta\}, \quad (1)$$

donde:

- $\Delta > 0$  es el radio de la región de confianza.
- $\|\cdot\|$  norma euclídeana en  $\mathbb{R}^n$
- $g = \nabla f(x)$
- $B$  matriz simétrica ( $B \approx \nabla^2 f(x)$ )

Lemma (Sorensen, 1983)

Si  $p$  es una solución de (1) entonces  $p$  es solución de la ecuación:

$$(B + \alpha I) p = -g, \quad (2)$$

con  $B + \alpha I$  semidefinida positiva,  $\alpha \geq 0$  y  $\alpha(\Delta - \|p\|) = 0$ .





# Clase de Algoritmos de Región de Confianza

Cuadrática restringida a una bola (elipse).

$$\min_w \{g^t w + \frac{1}{2} w^t B w : \|w\| \leq \Delta\}, \quad (1)$$

donde:

- $\Delta > 0$  es el radio de la región de confianza.
- $\|\cdot\|$  norma euclídeana en  $\mathbb{R}^n$
- $g = \nabla f(x)$
- $B$  matriz simétrica ( $B \approx \nabla^2 f(x)$ )

## Lemma (Sorensen, 1983)

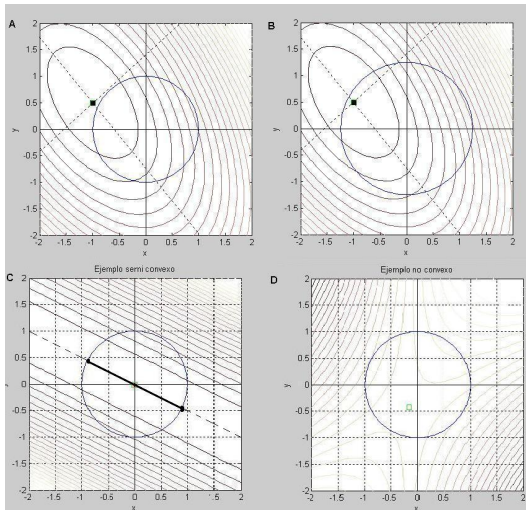
Si  $p$  es una solución de (1) entonces  $p$  es solución de la ecuación:

$$(B + \alpha I) p = -g, \quad (2)$$

con  $B + \alpha I$  semidefinida positiva,  $\alpha \geq 0$  y  $\alpha (\Delta - \|p\|) = 0$ .

# Región de Confianza

## Visión Geométrica



# Región de Confianza

## Hechos, tareas y metas

### Hechos

- $p$  depende de  $\alpha$  :  $p(\alpha) = -(B + \alpha I)^{-1}g$
- $B$  y  $B + \alpha I$  son simétricas (su espectro es real)
- Si  $\lambda_i$  es valor propio  $B$ , entonces  $\lambda_i + \alpha$  es valor propio  $B + \alpha I$ .
- $B + \alpha I$  es semidefinida positiva para  $\alpha \geq -\lambda_1$  donde  $\lambda_1$  es el menor valor propio de  $B$
- $\|p(\alpha)\|^2$  es una función racional.

### Tareas y Metas

- Resolver  $(B + \alpha I)p = -g$  eficientemente para distintos valores de  $\alpha$ .
- Construir una sucesión de actualizaciones para  $\alpha$
- Parar el proceso cuando  $\|p(\alpha)\| \approx \Delta$



# Región de Confianza

## Hechos, tareas y metas

### Hechos

- $p$  depende de  $\alpha$  :  $p(\alpha) = -(B + \alpha I)^{-1}g$
- $B$  y  $B + \alpha I$  son simétricas (su espectro es real)
- Si  $\lambda_i$  es valor propio  $B$ , entonces  $\lambda_i + \alpha$  es valor propio  $B + \alpha I$ .
- $B + \alpha I$  es semidefinida positiva para  $\alpha \geq -\lambda_1$  donde  $\lambda_1$  es el menor valor propio de  $B$
- $\|p(\alpha)\|^2$  es una función racional.

### Tareas y Metas

- Resolver  $(B + \alpha I)p = -g$  eficientemente para distintos valores de  $\alpha$ .
- Construir una sucesión de actualizaciones para  $\alpha$
- Parar el proceso cuando  $\|p(\alpha)\| \approx \Delta$



# Región de Confianza

## Hechos, tareas y metas

### Hechos

- $p$  depende de  $\alpha$  :  $p(\alpha) = -(B + \alpha I)^{-1}g$
- $B$  y  $B + \alpha I$  son simétricas (su espectro es real)
- Si  $\lambda_i$  es valor propio  $B$ , entonces  $\lambda_i + \alpha$  es valor propio  $B + \alpha I$ .
- $B + \alpha I$  es semidefinida positiva para  $\alpha \geq -\lambda_1$  donde  $\lambda_1$  es el menor valor propio de  $B$
- $\|p(\alpha)\|^2$  es una función racional.

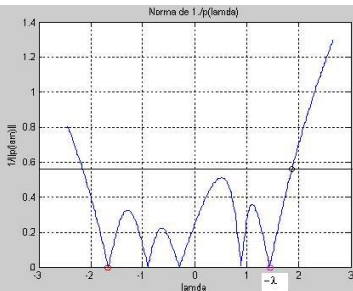
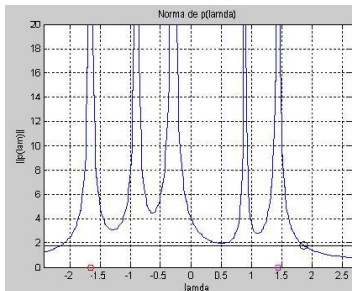
### Tareas y Metas

- Resolver  $(B + \alpha I)p = -g$  eficientemente para distintos valores de  $\alpha$ .
- Construir una sucesión de actualizaciones para  $\alpha$
- Parar el proceso cuando  $\|p(\alpha)\| \approx \Delta$



# Estabilidad de la ecuación $\|p(\alpha)\|^2 - \Delta^2 = 0$ .

$$\|p(\alpha)\|^2 = \sum_{i=1}^n \frac{\hat{g}_i^2}{(\lambda_i + \alpha)^2} \text{ tiene polos en } -\lambda_i \quad \frac{1}{\|p(\alpha)\|^2}$$



Uy!! muy fea,...

esta está mejor

$$\Phi(\alpha) \equiv \frac{1}{\Delta} - \frac{1}{\|p(\alpha)\|} = 0 \quad \longrightarrow \quad \alpha_+ = \alpha + \left( \frac{\|p\|}{\|q\|} \right)^2 \left( \frac{\|p\| - \Delta}{\Delta} \right)$$

$$R^t R p = -g; \quad R^t q = p \quad \text{donde } R \text{ es el factor de Cholesky de } \mathbf{B} + \alpha \mathbf{I} = R^t R$$



## Caso difícil: Algoritmo de Cline, Moler, Stewart y Wilkinson. En la foto aparecen Wilkinson y Givens



# Algoritmo Optimización tipo región de confianza

Sean  $0 < \mu < \eta < 1$  y  $0 < \gamma_1 < \gamma_2 < 1 < \gamma_3$

- Dados  $x_0 \in \mathbb{R}^n$ ,  $B_0$  simétrica y  $\Delta_0 > 0$
- For  $k=0, 1, 2, \dots$ : "hasta convergencia"
  - 1 Calcular  $g_k = \nabla f(x_k)$  y  $H_k \approx \nabla^2 f(x_k)$
  - 2 **Encontrar solución  $p_k$  aproximada de**  

$$\left\langle \min_p \left\{ g^t p + \frac{1}{2} p^t H_k p : \|p\| \leq \Delta \right\} \right\rangle$$
  - 3 Calcular  $\rho_k = \frac{(f(x_k + p_k) - f(x_k))}{g^t p_k + \frac{1}{2} p_k^t H_k p_k}$
  - 4 Si  $\rho_k \leq \mu$  entonces  $\Delta_k := \Delta \in [\gamma_1 \Delta_k; \gamma_2 \Delta_k]$  e ir a (2)
  - 5  $x_{k+1} := x_k + p_k$
  - 6 Si  $\rho_k \leq \eta$  entonces  $\Delta_{k+1} \in [\gamma_2 \Delta_k; \Delta_k]$  si no  $\Delta_{k+1} \in [\Delta_k; \gamma_3 \Delta_k]$



# Nuestra Contribución

- Nueva forma de factorizar  $B + \alpha I = R^T R$  para distintos valores de  $\alpha$
- Evaluación eficiente de  $q(p) = \frac{1}{2} p^T H p + g^T p$   
costo bruto:  $o(n^2 + 2n)$  flops  
se reduce a:  $o(n)$  flops
- Demostraciones y códigos no usuales

$$\bullet \quad \Phi(\alpha) \equiv \frac{1}{\Delta} - \frac{1}{\|p(\alpha)\|} = 0 \quad \rightarrow \quad \alpha_+ = \alpha + \left( \frac{\|p\|}{\|q\|} \right)^2 \left( \frac{\|p\| - \Delta}{\Delta} \right)$$

- Código del algoritmo de Cline, Moler, Stewart y Wilkinson.
- **Toolbox de MatLab** para resolver mínimos cuadrados no lineales mediante DGW, Huschens o Zhang + región de confianza.
- **Experimentos** comparando los resultados reportados por Zhang et al. [2003] realizados con búsqueda en línea con la regla inexacta de Armijo con el de región de confianza implementado en este trabajo.



# Nuestra Contribución

- Nueva forma de factorizar  $B + \alpha I = R^t R$  para distintos valores de  $\alpha$
- Evaluación eficiente de  $q(p) = \frac{1}{2} p^t H p + g^t p$   
costo bruto:  $o(n^2 + 2n)$  flops  
se reduce a:  $o(n)$  flops
- Demostraciones y códigos no usuales

$$\bullet \quad \Phi(\alpha) \equiv \frac{1}{\Delta} - \frac{1}{\|p(\alpha)\|} = 0 \quad \rightarrow \quad \alpha_+ = \alpha + \left( \frac{\|p\|}{\|q\|} \right)^2 \left( \frac{\|p\| - \Delta}{\Delta} \right)$$

- Código del algoritmo de Cline, Moler, Stewart y Wilkinson.
- **Toolbox de MatLab** para resolver mínimos cuadrados no lineales mediante DGW, Huschens o Zhang + región de confianza.
- **Experimentos** comparando los resultados reportados por Zhang et al. [2003] realizados con búsqueda en línea con la regla inexacta de Armijo con el de región de confianza implementado en este trabajo.



# Nuestra Contribución

- Nueva forma de factorizar  $B + \alpha I = R^t R$  para distintos valores de  $\alpha$
- Evaluación eficiente de  $q(p) = \frac{1}{2} p^t H p + g^t p$   
costo bruto:  $o(n^2 + 2n)$  flops  
se reduce a:  $o(n)$  flops
- Demostraciones y códigos no usuales

- $$\Phi(\alpha) \equiv \frac{1}{\Delta} - \frac{1}{\|p(\alpha)\|} = 0 \rightarrow \alpha_+ = \alpha + \left( \frac{\|p\|}{\|q\|} \right)^2 \left( \frac{\|p\| - \Delta}{\Delta} \right)$$
- Código del algoritmo de Cline, Moler, Stewart y Wilkinson.

- **Toolbox de MatLab** para resolver mínimos cuadrados no lineales mediante DGW, Huschens o Zhang + región de confianza.
- **Experimentos** comparando los resultados reportados por Zhang et al. [2003] realizados con búsqueda en línea con la regla inexacta de Armijo con el de región de confianza implementado en este trabajo.



# Nuestra Contribución

- Nueva forma de factorizar  $B + \alpha I = R^t R$  para distintos valores de  $\alpha$
- Evaluación eficiente de  $q(p) = \frac{1}{2} p^t H p + g^t p$   
costo bruto:  $o(n^2 + 2n)$  flops  
se reduce a:  $o(n)$  flops
- Demostraciones y códigos no usuales

- $$\Phi(\alpha) \equiv \frac{1}{\Delta} - \frac{1}{\|p(\alpha)\|} = 0 \rightarrow \alpha_+ = \alpha + \left( \frac{\|p\|}{\|q\|} \right)^2 \left( \frac{\|p\| - \Delta}{\Delta} \right)$$
- Código del algoritmo de Cline, Moler, Stewart y Wilkinson.

- **Toolbox de MatLab** para resolver mínimos cuadrados no lineales mediante DGW, Huschens o Zhang + región de confianza.
- **Experimentos** comparando los resultados reportados por Zhang et al. [2003] realizados con búsqueda en línea con la regla inexacta de Armijo con el de región de confianza implementado en este trabajo.



# Nuestra Contribución

- Nueva forma de factorizar  $B + \alpha I = R^t R$  para distintos valores de  $\alpha$
- Evaluación eficiente de  $q(p) = \frac{1}{2} p^t H p + g^t p$   
costo bruto:  $o(n^2 + 2n)$  flops  
se reduce a:  $o(n)$  flops
- Demostraciones y códigos no usuales

- $$\Phi(\alpha) \equiv \frac{1}{\Delta} - \frac{1}{\|p(\alpha)\|} = 0 \rightarrow \alpha_+ = \alpha + \left( \frac{\|p\|}{\|q\|} \right)^2 \left( \frac{\|p\| - \Delta}{\Delta} \right)$$
- Código del algoritmo de Cline, Moler, Stewart y Wilkinson.

- **Toolbox de MatLab** para resolver mínimos cuadrados no lineales mediante DGW, Huschens o Zhang + región de confianza.
- **Experimentos** comparando los resultados reportados por Zhang et al. [2003] realizados con búsqueda en línea con la regla inexacta de Armijo con el de región de confianza implementado en este trabajo.



## Ejemplos de experimentos numéricos

Función Helical Valley: cero-residual pero fuertemente no lineal

HELICAL VALLEY FUNCTION				
Residual Cero pero fuertemente no lineal				
	TSSM		GAUSS-NEWTON	
<i>iter</i>	$\ \nabla f(x)\ $	$\ e_i\ / e_{i-1} $	$\ \nabla f(x)\ $	$\ e_i\ / e_{i-1} $
0	0.69009e+003	-----	0.69009e+003	-----
1	0.28090e+003	0.4056e-000	0.35890e+003	0.5499e-000
2	0.82000e+002	0.2421e-000	0.25900e+003	0.7058e-000
3	0.26970e+001	0.1388e-000	0.15090e+003	0.4981e-000
4	0.18150e-000	0.3315e-000	0.84600e+002	0.4683e-000
5	0.12320e-001	0.4239e-001	0.34790e+002	0.2559e-000
6	0.29250e-004	0.4805e-002	0.31960e+001	0.8625e-001
7	0.52760e-009	0.1117e-004	0.26860e-001	0.7678e-002
8			0.14770e-005	0.5929e-004
9			0.54430e-014	0.2979e-008

# Ejemplos de experimentos numéricos

Función paramétrica : Para  $\lambda = 1$  es cero-residual.

FUNCIÓN PARAMÉTRICA				
$\lambda=1$ . Residual Cero				
	TSSM		SSM	
iter	$\ \nabla f(x)\ $	$\ e_i\ / e_{i-1} $	$\ \nabla f(x)\ $	$\ e_i\ / e_{i-1} $
0	0.27400e+001	-----	0.27400e+001	-----
1	0.12086e+001	0.67710e-000	0.12080e+001	0.67710e-000
2	0.35730e-000	0.39770e-000	0.39110e-000	0.42290e-000
3	0.80790e-001	0.16850e-000	0.17220e-000	0.30090e-000
4	0.35510e-002	0.42530e-001	0.68960e-001	0.40720e-000
5	0.10660e-004	0.30100e-002	0.85500e-002	0.13170e-000
6	0.57880e-010	0.54260e-005	0.77210e-004	0.90980e-002
7			0.51660e-006	0.66910e-002
8			0.11010e-009	0.21320e-003

## Ejemplos de experimentos numéricos

Función paramétrica : Para  $\lambda \neq 1$  es no cero-residual ( $\lambda = 0.5$ )

FUNCIÓN PARAMÉTRICA				
$\lambda=0.5$ . Residual Distinto de Cero				
	TSSM		SSM (BFGS)	
<i>iter</i>	$\ \nabla f(x)\ $	$\ e_i\ / e_{i-1} $	$\ \nabla f(x)\ $	$\ e_i\ / e_{i-1} $
0	0.2140e+001	-----	0.2140e+001	-----
1	0.1139e+001	0.5791e-000	0.1139e+001	0.5791e-000
2	0.4460e-000	0.2294e-000	0.3785e-000	0.1833e-000
3	0.4845e-001	0.2963e-000	0.9498e-001	0.4902e-000
4	0.1426e-001	0.2873e-000	0.2472e-001	0.3846e-000
5	0.6981e-004	0.3128e-002	0.6679e-002	0.1641e-000
6	0.1788e-005	0.1249e-001	0.6984e-003	0.5462e-001
7	0.6495e-007	0.3749e-001	0.2146e-004	0.4512e-001
8	0.2415e-008	0.1782e-003	0.3171e-006	0.2933e-001
9			0.9056e-008	0.2730e-001
10			0.3437e-009	0.9518e-002



# Comparación con resultados de Zhang et al.

## Problemas cero-residuales

No.	Método	Búsqueda en línea			Región de Confianza			
		f eval	Iter	tpo	Fac	Iter	$\ g\ $	tpo
1	DGW	152	71	3.24	90	70	4.0168e-008	0.79
	Hus	150	70	3.13	93	71	2.4825e-014	0.49
	Zhang	150	70	3.13	93	71	2.4825e-014	0.51
2	DGW	165	63	37.35	130	93	8.9238e-006	10.12
	Hus	149	55	33.23	68	44	1.7295e-009	6.91
	Zhang	149	55	33.17	68	44	1.7296e-009	6.88
3	DGW	37	18	1.54	20	19	3.4637e-008	0.13
	Hus	29	14	1.21	15	14	3.6447e-008	0.08
	Zhang	29	14	1.21	15	14	3.6447e-008	0.08
4	DGW	11	5	0.22	6	5	2.4385e-013	0.30
	Hus	10	5	0.16	5	4	2.6678e-009	0.27
	Zhang	10	5	0.22	5	4	2.6678e-009	0.28
5	DGW	14	7	0.16	8	7	6.1654e-012	0.18
	Hus	13	6	0.16	7	6	2.2198e-011	0.13
	Zhang	11	5	0.17	7	6	1.8739e-012	0.13

# Comparación con resultados de Zhang et al.

Problemas cero-residuales...continuación

		Búsqueda en línea			Región de Confianza			
No.	Método	f eval	Iter	tpo	Fac	Iter	$\ g\ $	tpo
6	DGW	925	414	106.01	29	22	6.3183e-008	3.94
	Hus	215	58	23.51	23	15	8.7641e-008	3.05
	Zhang	210	65	23.28	30	16	8.4127e-008	3.86
7	DGW	584	258	28.45	348	267	7.0173e-009	9.55
	Hus	580	256	28.51	342	266	2.3104e-012	9.58
	Zhang	544	252	27.29	338	265	2.7005e-013	9.39
8	DGW	63	27	3.02	31	10	6.1287e+000	4.35
	Hus	62	26	2.97	30	9	6.2098e+000	4.13
	Zhang	60	26	2.85	30	9	8.2918e+000	9.58
9	DGW	50	25	0.61	28	6	8.4834e-002	0.81
	Hus	54	27	0.66	26	5	2.6665e+000	0.55
	Zhang	75	20	0.82	44	17	5.9502e-001	1.17
10	DGW	26	12	2.81	22	17	1.7567e-008	7.72
	Hus	24	11	2.64	32	24	1.1178e-010	9.34
	Zhang	17	8	1.87	34	26	1.7049e-010	13.56

# Comparación con resultados de Zhang et al.

## Problemas no cero-residuales

No.	Método	Búsqueda en línea			Región de Confianza			
		f eval	Iter	tpo	Fac	Iter	$\ g\ $	tpo
1	DGW	118	57	21.14	36	12	5.3021e-002	2.82
	Hus	118	57	21.04	48	19	1.0062e-003	4.13
	Zhang	81	37	14.11	32	9	1.9426e+000	2.24
2	DGW	35	15	4.34	36	15	6.6830e-006	5.08
	Hus	35	15	4.12	16	14	5.6134e-006	2.96
	Zhang	26	13	3.46	15	13	1.4125e-005	2.90
3	DGW	50	25	7.14	74	45	1.9123e-004	2.68
	Hus	50	25	7.08	62	34	6.3948e-005	1.86
	Zhang	40	20	5.82	62	33	2.5012e-004	1.95
4	DGW	106	36	7.34	40	19	1.8124e-005	4.96
	Hus	106	36	7.69	46	22	2.0158e-005	5.21
	Zhang	91	29	6.59	42	21	2.5233e-005	5.00
5	DGW	162	54	19.06	52	28	6.6666e-005	2.94
	Hus	162	54	18.53	56	29	1.7508e-005	2.90
	Zhang	110	36	12.86	96	53	3.3596e-005	4.63
6	DGW	143	43	10.05	25	4	1.8224e+002	0.24
	Hus	143	43	10.11	24	4	1.6343e+002	0.22
	Zhang	143	43	10.05	24	4	1.8555e+002	0.23

# Comparación con resultados de Zhang et al.

## Problemas no cero-residuales...continuación

		Búsqueda en línea			Región de Confianza			
No.	Método	f eval	Iter	tpo	Fac	Iter	$\ g\ $	tpo
7	DGW	26	10	7.03	129	103	1.2761e-002	69.82
	Hus	26	10	6.98	136	109	1.5233e-002	82.79
	Zhang	18	8	5.16	18	17	3.6634e+002	16.12
8	DGW	10	5	0.11	6	5	3.2310e-008	0.08
	Hus	10	5	0.11	6	5	4.6683e-010	0.02
	Zhang	10	5	0.16	6	5	1.0937e-009	0.04
9	DGW	49	19	1.97	35	22	2.7481e-010	0.23
	Hus	41	15	1.70	32	22	4.7182e-009	0.19
	Zhang	38	13	1.48	21	16	4.7357e-011	0.12
10	DGW	205	77	5.88	47	33	6.5179e-006	1.42
	Hus	166	65	4.56	111	69	1.8910e-006	7.41
	Zhang	119	44	3.24	243	216	1.0415e-004	22.67
11	DGW	108	54	9.00	24	3	1.8476e+002	0.25
	Hus	108	54	8.95	24	3	1.3698e+002	0.22
	Zhang	86	43	7.09	24	4	1.2975e+002	0.20
12	DGW	106	53	1.26	122	95	8.3056e-003	0.49
	Hus	106	53	1.16	169	112	3.0777e-004	0.74
	Zhang	55	27	0.66	182	116	2.0857e-004	0.98



# Factorización de $B + \lambda I$

Transformación a un problema de mínimos cuadrados lineales

Para cada nuevo  $\lambda$  hay que refactorizar  $B + \lambda I = R^t R$  para resolver  
 $(B + \lambda I)p = -g$ ,  
 ¿cómo hacerlo eficientemente, si ya existe  $B = R^t R$ ?

$$\min \left\| \begin{bmatrix} R \\ \sqrt{\lambda} I \end{bmatrix} p + \begin{bmatrix} R^{-t} g \\ 0 \end{bmatrix} \right\|^2$$

$$\Downarrow$$

$$\begin{bmatrix} R^t & \sqrt{\lambda} I \end{bmatrix} \begin{bmatrix} R \\ \sqrt{\lambda} I \end{bmatrix} p = - \begin{bmatrix} R^t & \sqrt{\lambda} I \end{bmatrix} \begin{bmatrix} R^{-t} g \\ 0 \end{bmatrix}$$

$$\Downarrow$$

$$(R^t R + \lambda I) p = -g = (B + \lambda I) p$$



# Factorización de $B + \lambda I$

Transformación a un problema de mínimos cuadrados lineales

Para cada nuevo  $\lambda$  hay que refactorizar  $B + \lambda I = R^t R$  para resolver  
 $(B + \lambda I)p = -g$ ,  
 ¿cómo hacerlo eficientemente, si ya existe  $B = R^t R$ ?

$$\min \left\| \begin{bmatrix} R \\ \sqrt{\lambda} I \end{bmatrix} p + \begin{bmatrix} R^{-t} g \\ 0 \end{bmatrix} \right\|^2$$

$$\Downarrow$$

$$\begin{bmatrix} R^t & \sqrt{\lambda} I \end{bmatrix} \begin{bmatrix} R \\ \sqrt{\lambda} I \end{bmatrix} p = - \begin{bmatrix} R^t & \sqrt{\lambda} I \end{bmatrix} \begin{bmatrix} R^{-t} g \\ 0 \end{bmatrix}$$

$$\Downarrow$$

$$(R^t R + \lambda I) p = -g = (B + \lambda I) p$$



# Factorización de $B + \lambda I$

Transformación a un problema de mínimos cuadrados lineales

Para cada nuevo  $\lambda$  hay que refactorizar  $B + \lambda I = R^t R$  para resolver  
 $(B + \lambda I)p = -g$ ,  
 ¿cómo hacerlo eficientemente, si ya existe  $B = R^t R$ ?

$$\begin{aligned} \min \left\| \begin{bmatrix} R \\ \sqrt{\lambda} I \end{bmatrix} p + \begin{bmatrix} R^{-t} g \\ 0 \end{bmatrix} \right\|^2 \\ \Downarrow \\ \begin{bmatrix} R^t & \sqrt{\lambda} I \end{bmatrix} \begin{bmatrix} R \\ \sqrt{\lambda} I \end{bmatrix} p = - \begin{bmatrix} R^t & \sqrt{\lambda} I \end{bmatrix} \begin{bmatrix} R^{-t} g \\ 0 \end{bmatrix} \\ \Downarrow \\ (R^t R + \lambda I) p = -g = (B + \lambda I) p \end{aligned}$$



# Factorización de $B + \lambda I$

## Factorización QR vía rotaciones de Givens

$$\text{Estructura de } M \equiv \begin{bmatrix} R \\ \sqrt{\lambda} I \end{bmatrix} = \begin{bmatrix} r_{11} & \dots & r_{1n} \\ 0 & \ddots & \vdots \\ 0 & 0 & r_{nn} \\ \sqrt{\lambda} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sqrt{\lambda} \end{bmatrix} \quad M \in \mathbb{R}^{2n \times n}$$

$$M = \begin{bmatrix} r_{11} & \dots & r_{1n} \\ 0 & \ddots & \vdots \\ 0 & 0 & r_{nn} \\ \sqrt{\lambda} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sqrt{\lambda} \end{bmatrix} = \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \times & Q & \times \\ \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{bmatrix} \begin{bmatrix} \check{r}_{11} & \dots & \check{r}_{1n} \\ 0 & \ddots & \vdots \\ 0 & 0 & \check{r}_{nn} \end{bmatrix}$$

Vía rotaciones de Givens (No reflexiones de Householder)





# Factorización de $B + \alpha I$

¿Cuándo se puede aplicar la idea anterior?

Supongamos  $\alpha \in (-\lambda_1; \alpha^*)$  donde  $\phi(\alpha^*) = 0$  y tenemos la factorización de Cholesky de

$$\tilde{B} = B + \alpha I = R^t R$$

y necesitamos la factorización

$$B + \alpha_+ I = R_+^t R_+,$$

entonces

$$B + \alpha_+ I = B + \alpha I + (\alpha_+ - \alpha) I = \tilde{B} + \tilde{\alpha} I$$

si  $\tilde{\alpha} = \alpha_+ - \alpha > 0$  podemos aplicar la idea anterior.

*En otras palabras: la sucesión  $\{\alpha_k\}$  debe ser creciente!!*



# Factorización de $B + \alpha I$

¿Cuándo se puede aplicar la idea anterior?

Supongamos  $\alpha \in (-\lambda_1; \alpha^*)$  donde  $\phi(\alpha^*) = 0$  y tenemos la factorización de Cholesky de

$$\tilde{B} = B + \alpha I = R^t R$$

y necesitamos la factorización

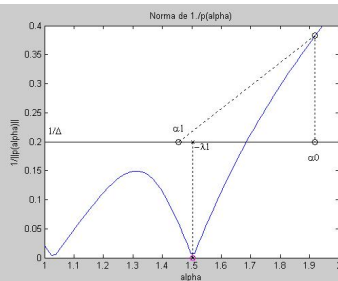
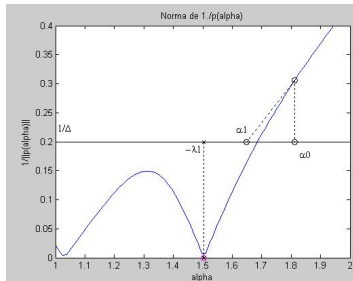
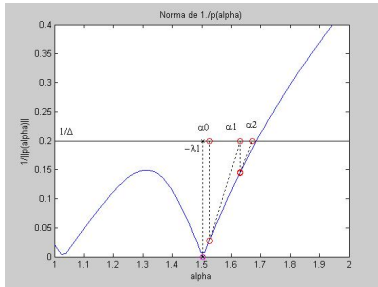
$$B + \alpha_+ I = R_+^t R_+,$$

entonces

$$B + \alpha_+ I = B + \alpha I + (\alpha_+ - \alpha) I = \tilde{B} + \tilde{\alpha} I$$

si  $\tilde{\alpha} = \alpha_+ - \alpha > 0$  podemos aplicar la idea anterior.  
*En otras palabras: la sucesión  $\{\alpha_k\}$  debe ser creciente!!*





**MUCHAS  
GRACIAS**

