

Coloquio de Matemática Aplicada

Modelo sustituto basado en codificación dispersa
para resolver problemas costosos de optimización
multiobjetivo

Cindy Calderón-Arce

Escuela de Matemática

Enero, 2021

Contenidos

Preliminares
Antecedentes
Propuesta
Resultados
Conclusiones

- 1 Preliminares
 - Contexto
 - Problema
- 2 Antecedentes
 - Fundamentación
- 3 Propuesta
- 4 Resultados
- 5 Conclusiones

Contexto

Problemas reales



Problemas de optimización



Multiobjetivo

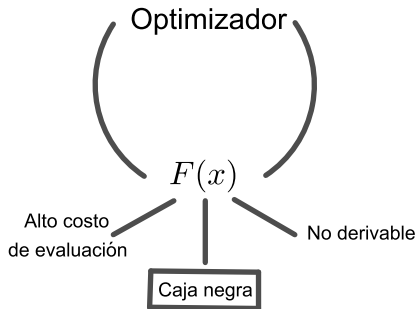


Funciones objetivo de alto costo de evaluación



Contexto

Optimización multiobjetivo costosa



Preliminares

Contexto

Problema

Antecedentes

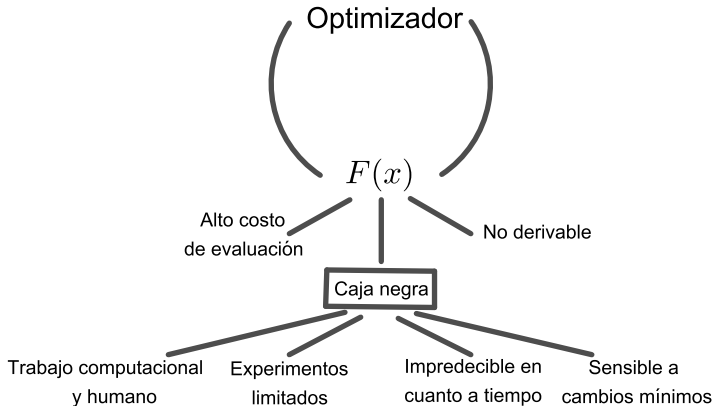
Propuesta

Resultados

Conclusiones

Contexto

Optimización multiobjetivo costosa



Problema

Reducir el consumo de recursos al resolver problemas de optimización **multiobjetivo** definidos a partir de funciones objetivo **complejas** y de **alto costo** de evaluación.

$$\text{mín } F(x) = (f_1(x), f_2(x), f_3(x), \dots, f_K(x))$$

$$\text{con } f_k : \mathbb{R}^n \rightarrow \mathbb{R}.$$

Problema

Preliminares

Contexto

Problema

Antecedentes

Propuesta

Resultados

Conclusiones

Reducir el consumo de recursos al resolver problemas de optimización **multiobjetivo** definidos a partir de funciones objetivo **complejas** y de **alto costo** de evaluación.

$$\text{mín } F(x) = (f_1(x), f_2(x), f_3(x), \dots, f_K(x))$$

con $f_k : \mathbb{R}^n \rightarrow \mathbb{R}$.

Antecedentes

Optimización
multiobjetivo

- Métodos clásicos
- Escalarización
- Heurísticas (Algoritmos genéticos)
- Modelos sustitutos

Codificación dispersa

- Reconstrucción de señales
- Descenso gradiente (GD)

Preliminares

Antecedentes

Fundamentación

Propuesta

Resultados

Conclusiones

Fundamentación

- 1 Heurísticas, diseñadas para resolver problemas de optimización multiobjetivo complejos, requieren de reiteradas llamadas o evaluaciones a las funciones objetivo.
- 2 Modelos sustitutos requieren de un número considerable de muestras, que dependen de la dimensión del espacio de búsqueda y de las características de la función a aproximar.

Fundamentación

- 3 Métodos para reconstruir o representar señales generalmente requieren de una cantidad exhaustiva de datos o muestras.
- 4 Codificación dispersa representa una señal a partir de pocas muestras y además transfiere lo aprendido a contextos similares.
- 5 No se han encontrado referencias del uso de codificación dispersa como técnica para la reconstrucción o representación matemática de una función en contextos de optimización.

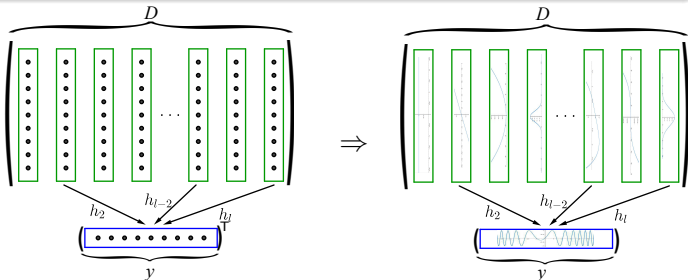
Propuesta

Codificación dispersa

$$y \approx Dh$$

$$D^*, h^* = \arg \min_{D, h} \{ \|y - Dh\|_2^2 + \lambda \|h\|_1 \}$$

D : diccionario – h : vector disperso



Propuesta

Diccionario paramétrico

$$D(x) = (\phi_1(x), \phi_2(x), \phi_3(x), \dots, \phi_l(x))$$

con

$$\begin{aligned} \phi_j(x) &= \exp \left(-\frac{1}{2} (x - \mu^{(j)})^T (\Sigma^{(j)})^{-1} (x - \mu^{(j)}) \right) \\ &= \exp \left(-\frac{1}{2} \left(\sum_{i=1}^n (x_i - \mu_i^{(j)})^2 \rho_{i,i}^{(j)} + \dots \right. \right. \\ &\quad \left. \left. 2 \sum_{i=1}^{n-1} \sum_{g=i+1}^n (x_i - \mu_i^{(j)})(x_g - \mu_g^{(j)}) \rho_{i,g}^{(j)} \right) \right) \end{aligned}$$

Propuesta

Diccionario paramétrico

$$D(x)^*, h^* = \arg \min_{D, h} \{ \|f(x) - D(x)h\|_{2}^2 + \lambda \|h\|_1 \}$$



$$D = \begin{pmatrix} D(x^{(1)}) \\ D(x^{(2)}) \\ \vdots \\ D(x^{(m)}) \end{pmatrix} = \begin{pmatrix} \phi_1(x^{(1)}) & \phi_2(x^{(1)}) & \phi_3(x^{(1)}) & \dots & \phi_l(x^{(1)}) \\ \phi_1(x^{(2)}) & \phi_2(x^{(2)}) & \phi_3(x^{(2)}) & \dots & \phi_l(x^{(2)}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_1(x^{(m)}) & \phi_2(x^{(m)}) & \phi_3(x^{(m)}) & \dots & \phi_l(x^{(m)}) \end{pmatrix}$$

$X = \{x^{(1)}, x^{(2)}, \dots, x^{(m)}\}$ es el conjunto de datos observados y $f(X) = \{f(x^{(1)}), f(x^{(2)}), \dots, f(x^{(m)})\}$ sus correspondientes valores objetivo.

Propuesta

Diccionario paramétrico

$$D(x)^*, h^* = \arg \min_{D, h} \{ \|f(x) - D(x)h\|_{2}^2 + \lambda \|h\|_1 \}$$

$$\Downarrow$$

$$D = \begin{pmatrix} D(x^{(1)}) \\ D(x^{(2)}) \\ \vdots \\ D(x^{(m)}) \end{pmatrix} = \begin{pmatrix} \phi_1(x^{(1)}) & \phi_2(x^{(1)}) & \phi_3(x^{(1)}) & \dots & \phi_l(x^{(1)}) \\ \phi_1(x^{(2)}) & \phi_2(x^{(2)}) & \phi_3(x^{(2)}) & \dots & \phi_l(x^{(2)}) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \phi_1(x^{(m)}) & \phi_2(x^{(m)}) & \phi_3(x^{(m)}) & \dots & \phi_l(x^{(m)}) \end{pmatrix}$$

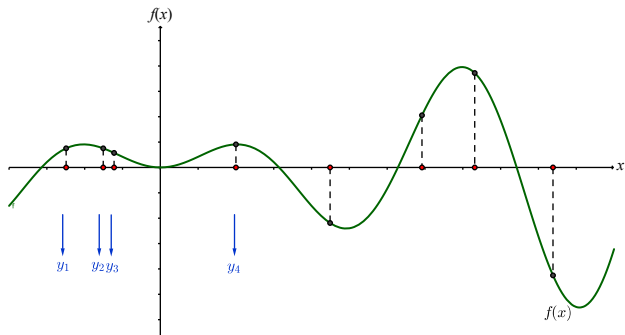
$X = \{x^{(1)}, x^{(2)}, \dots, x^{(m)}\}$ es el conjunto de datos observados y $f(X) = \{f(x^{(1)}), f(x^{(2)}), \dots, f(x^{(m)})\}$ sus correspondientes valores objetivo.

Propuesta

Conjunto de entrenamiento

$$y = \left(f(x^{(1)}), f(x^{(2)}), f(x^{(3)}), \dots, f(x^{(m)}) \right)^T$$

donde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ y cada $x^{(r)} \in \mathbb{R}^n$



Propuesta

SESM: Sparse encoded based surrogate model

$$D(x)^*, h^* = \arg \min_{D, h} \{ \|y - Dh\|_2^2 + \lambda \|h\|_1 \}$$

D es constante respecto a x pero no respecto a los parámetros de codificación $\mu^{(j)}$ y $(\Sigma^{(j)})^{-1}$ de cada función base ϕ_j . Dichos parámetros de codificación se almacenan en una matriz de la forma

$$\Theta = (\theta_1, \theta_2, \theta_3, \dots, \theta_l)$$

donde θ_j contiene que los parámetros de condificación de ϕ_j .

- Preliminares
- Antecedentes
- Propuesta
- Resultados
- Conclusiones

Propuesta

- h se ajusta por medio de ISTA (Iterative Shrinkage and Thresholding Algorithm)
- Θ se ajusta por medio de GD que minimiza el error cuadrático medio entre $f(X)$ y $\psi(X)$.

SESM: Sparse encoded based surrogate model

$$\Theta_{p,j} = \Theta_{p,j} - \alpha \frac{\partial E(\Theta)}{\partial \Theta_{p,j}}$$

con

$$E(\Theta) = \frac{1}{2m} \sum_{r=1}^m \left(\psi(x^{(r)}) - f(x^{(r)}) \right)^2 = \frac{1}{2m} \sum_{r=1}^m \left(\psi^{(r)} - y^{(r)} \right)^2$$

Propuesta

- h se ajusta por medio de ISTA (Iterative Shrinkage and Thresholding Algorithm)
- Θ se ajusta por medio de GD que minimiza el error cuadrático medio entre $f(X)$ y $\psi(X)$.

SESM: Sparse encoded based surrogate model

$$\Theta_{p,j} = \Theta_{p,j} - \alpha \frac{\partial E(\Theta)}{\partial \Theta_{p,j}}$$

con

$$E(\Theta) = \frac{1}{2m} \sum_{r=1}^m \left(\psi(x^{(r)}) - f(x^{(r)}) \right)^2 = \frac{1}{2m} \sum_{r=1}^m \left(\psi^{(r)} - y^{(r)} \right)^2$$

Propuesta

- Preliminares
- Antecedentes
- Propuesta
- Resultados
- Conclusiones

SESM: Sparse encoded based surrogate model

$$\psi(x) := D(x) \cdot h = \sum_{j=1}^I \phi_j(x) h_j$$

donde

$$D(x) = (\phi_1(x), \phi_2(x), \phi_3(x), \dots, \phi_I(x))$$

Sparse encoded based surrogate model (SESM)

Preliminares

Antecedentes

Propuesta

Resultados

Conclusiones

Partición del espacio paramétrico

Por cada sub-bloque activo

$X^{(t_{ext})} \rightarrow$ Normalización

$Y^{(t_{ext})} \rightarrow$ Normalización

$D^{(t_{ext})}$

$\Theta \leftarrow$ GD

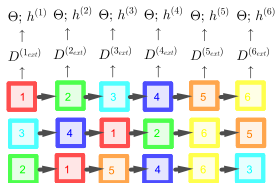
$h^{(t)} \leftarrow$ ISTA

X_{new}

$X = X \cup X_{new}$

Secuencial

Simultáneo



$$D = (D^{(1_{ext})} D^{(2_{ext})} D^{(3_{ext})} D^{(4_{ext})} D^{(5_{ext})} D^{(6_{ext})})$$

\Downarrow

$$\Theta; \mathcal{H} = (h^{(1)} h^{(2)} h^{(3)} h^{(4)} h^{(5)} h^{(6)})$$

Sparse encoded based surrogate model (SESM)

Preliminares
Antecedentes

Propuesta

Resultados

Conclusiones

Partición del espacio paramétrico

Por cada sub-bloque activo

$X^{(t_{ext})} \rightarrow$ Normalización

$Y^{(t_{ext})} \rightarrow$ Normalización

$D^{(t_{ext})}$

$\Theta \leftarrow$ GD

$h^{(t)} \leftarrow$ ISTA

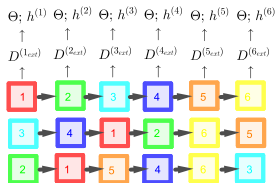
X_{new}

$X = X \cup X_{new}$

Mejorando el ajuste en los posibles mínimos

Secuencial

Simultáneo



$$D = (D^{(1_{ext})} \ D^{(2_{ext})} \ D^{(3_{ext})} \ D^{(4_{ext})} \ D^{(5_{ext})} \ D^{(6_{ext})})$$

\Downarrow

$$\Theta; \mathcal{H} = (h^{(1)} \ h^{(2)} \ h^{(3)} \ h^{(4)} \ h^{(5)} \ h^{(6)})$$

Propuesta

NSGAI-SESM

Modelo sustituto



Heurística de optimización



Codificación dispersa

Algoritmo genético



SESM



NSGA-II

Propuesta

NSGAI-SESM

Modelo sustituto



Heurística de optimización



Codificación dispersa



Algoritmo genético



SESM



NSGA-II



Propuesta

NSGAI-SESM

Modelo sustituto

 \Leftrightarrow

Heurística de optimización



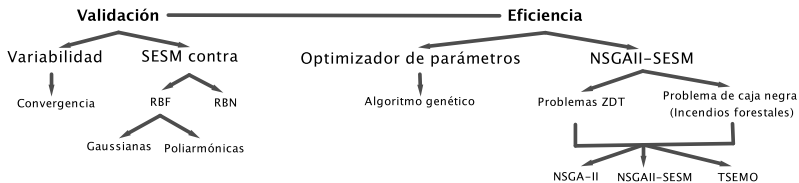
Codificación dispersa

Algoritmo genético

**SESM** \Leftrightarrow **NSGA-II**

Resultados

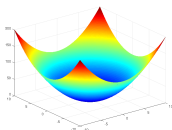
- Preliminares
- Antecedentes
- Propuesta
- Resultados**
- Conclusiones



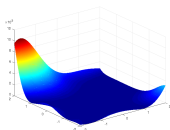
Resultados

Espacio de prueba

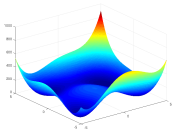
- Funciones definidas en \mathbb{R}^2



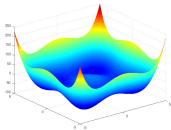
Sphere: SP



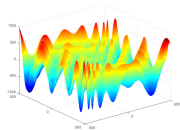
Goldstein Price: GP



Himmelblau: HB



Styblinski Tang: ST



Eggholder: EH

Resultados

Refinar la partición ayuda a la convergencia de SESM en el modo simultáneo, sin embargo, en el modo secuencial esto produce ruido en los resultados, alterando la convergencia del método en este modo.

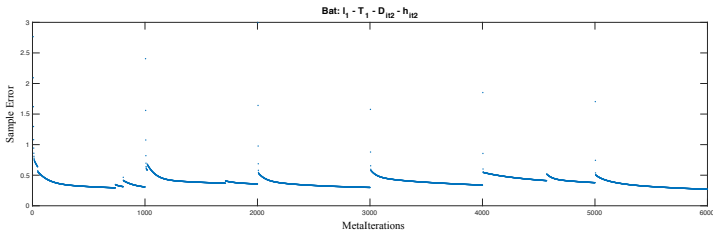


Fig. Errores obtenidos con SESM al aproximar la función GP con Bat: $l_1 - T_1 - D_{it2} - h_{it2}$

Resultados

Comparación con RBF

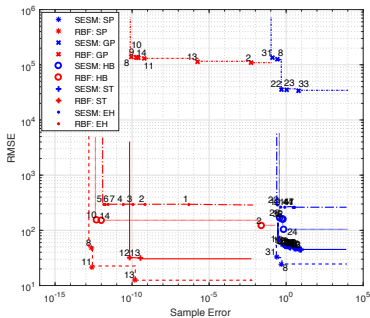
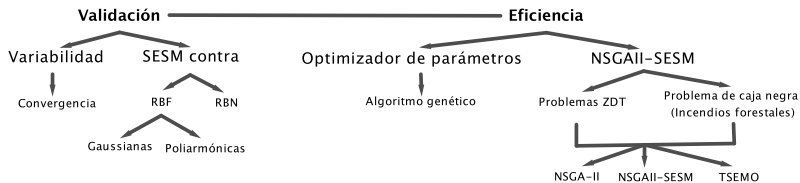


Fig. Frentes de Pareto de los errores obtenidos con cada función de prueba en \mathbb{R}^2

Resultados

- Preliminares
- Antecedentes
- Propuesta
- Resultados**
- Conclusiones



Resultados

Optimizador de parámetros

$$\underset{f}{\text{mín}} \quad \left(\underset{f}{\text{máx}} \{ \mathcal{R}(f) \}, \underset{f}{\text{máx}} \{ \mathcal{E}(f) \}, \underset{f}{\text{máx}} \{ \text{RMSE}(f) \}, \underset{f}{\text{máx}} \{ T(f) \} \right)$$

$$\text{s.t.} \quad f \in \{ SP, GP, HB, ST, EH \}$$

$$lb_T \leq T \leq ub_T$$

$$lb_l \leq l \leq ub_l$$

$$lb_h \leq D_{iter} \leq ub_h$$

$$lb_D \leq h_{iter} \leq ub_D$$

$$T, l, h_{iter}, D_{iter} \in \mathbb{Z}$$

Resultados

Optimizador de parámetros

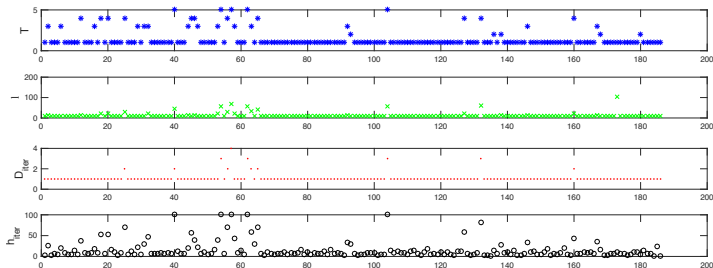


Fig. Parámetros de los puntos en el frente de Pareto resultante al optimizar los hiperparámetros de SESM

$I \in \{10, 20, \dots, 200\}$; $T \in \{1, 2, 3, 4, 5\}$; $D_{iter} \in \{1, 2, \dots, 100\}$; $h_{iter} \in \{1, 2, \dots, 100\}$

Resultados

Optimizador de parámetros

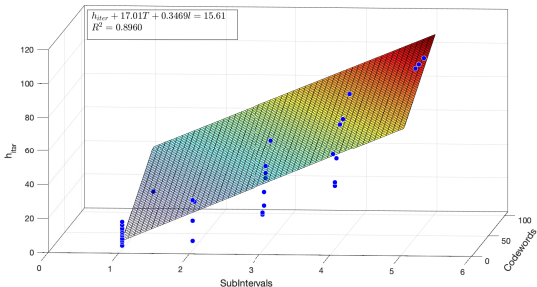
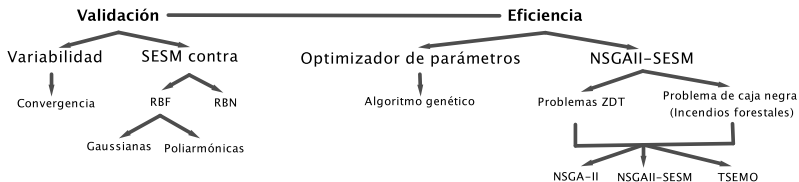


Fig. Ajuste del primer nivel de todas las configuraciones que generan el frente de Pareto resultante del optimizador de parámetros

Resultados

- Preliminares
- Antecedentes
- Propuesta
- Resultados**
- Conclusiones



Resultados

- Preliminares
- Antecedentes
- Propuesta
- Resultados**
- Conclusiones

NSGAI-SESM

- Problemas multiobjetivo ZDT, con un espacio de búsqueda de 10 dimensiones
- Comparación con
 - NSGA-II (Deb et. al, 2000)
 - TSEMO (Bradford, Schweidtmann y Lapkin, 2018)

Resultados

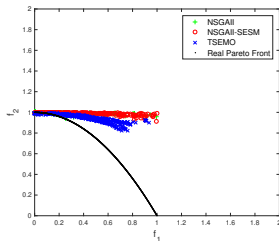
Estadísticos obtenidos al resolver 10 veces cada problema ZDT

Medida	Tiempo		Evaluaciones	
Método	NSGA-II			
Problema	Media	SD	Media	SD
ZDT1	0.3815	0.0728	2550	0
ZDT2	0.3447	0.0173	2550	0
ZDT3	0.3428	0.0081	2550	0
ZDT4	0.3410	0.0084	2550	0
ZDT6	0.3390	0.0087	2550	0
Método	NSGAI-SESM			
Problema	Media	SD	Media	SD
ZDT1	170.31	43.74	924.8	184.70
ZDT2	127.08	53.37	739.80	229.81
ZDT3	129.08	48.76	751.30	215.28
ZDT4	175.90	49.29	1056.60	274.24
ZDT6	156.81	72.35	804.20	250.90
Método	TSEMO			
Problema	Media	SD	Media	SD
ZDT1	3055.79	0.2582	1050	0
ZDT2	2960.19	0.5015	1050	0
ZDT3	2963.09	0.0834	1050	0
ZDT4	3059.14	0.3340	1050	0
ZDT6	2939.74	0.2709	1050	0

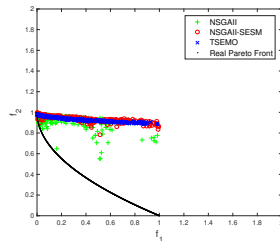
- Preliminares
- Antecedentes
- Propuesta
- Resultados**
- Conclusiones

Resultados

- Preliminares
- Antecedentes
- Propuesta
- Resultados
- Conclusiones



ZDT2



ZDT4

Fig. Frentes de Pareto obtenidos al resolver los problemas ZDT con NSGA-II, NSGAII-SESM y TSEMO

Resultados

- Preliminares
- Antecedentes
- Propuesta
- Resultados
- Conclusiones

Aplicación: Incendios forestales

Problema: Determinar la configuración del predictor que minimice el error y el tiempo de ejecución al determinar áreas quemadas en incendios forestales.

- Base de datos con las 517 muestras
- 12 atributos o valores de entrada

Resultados

- Preliminares
- Antecedentes
- Propuesta
- Resultados**
- Conclusiones

Aplicación: Incendios forestales

Problema: Determinar la configuración del predictor que minimice el error y el tiempo de ejecución al determinar áreas quemadas en incendios forestales.

- Base de datos con las 517 muestras
- 12 atributos o valores de entrada

Resultados

Aplicación: Incendios forestales

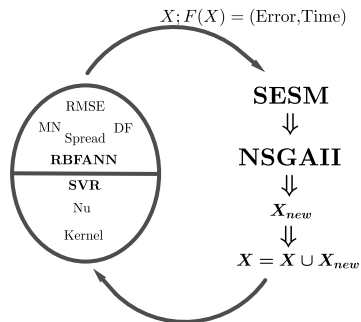
Predictores propuestos:

- RBFANN (RBN): red neuronal artificial de funciones de base radial.

Hiperparámetros considerados: error cuadrático medio (RMSE), propagación (Spread), número máximo de neuronas (MN), cantidad de neuronas nuevas por iteración (DF).

- SVR: regresor de soporte vectorial.

Hiperparámetros considerados: tipo de núcleo (Kernel), parámetro de regularización (Nu)



Resultados

Aplicación: Incendios forestales

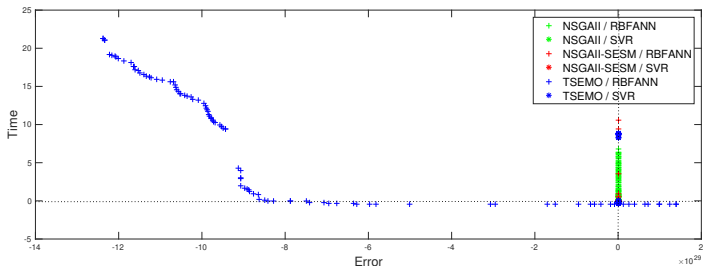


Fig. Frentes de Pareto determinados por NSGAII, NSGAII-SESM y TSEMO al ajustar RBFANN y SVR

Resultados

Aplicación: Incendios forestales

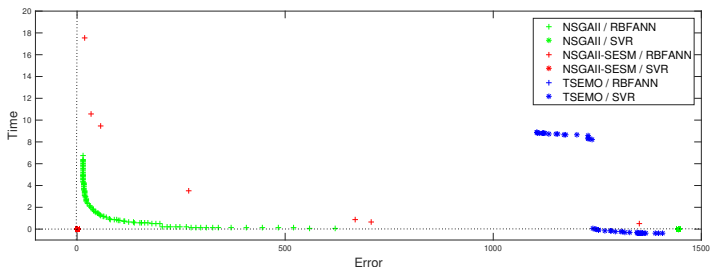


Fig. Región factible de frentes de Pareto determinados por NSGAII, NSGAII-SESM y TSEMO al ajustar RBFANN y SVR

Resultados

Aplicación: Incendios forestales

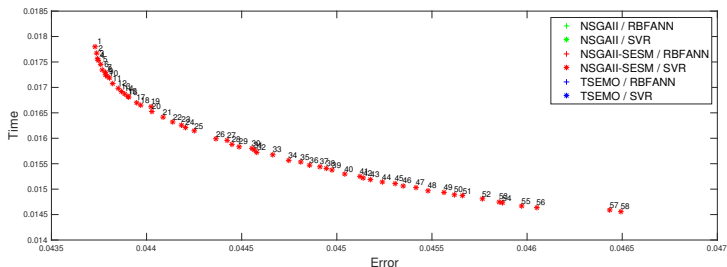


Fig. Acercamiento en frentes de Pareto determinados por NSGAI, NSGAI-SESM y TSEMO al ajustar RBFANN y SVR

Resultados

Aplicación: Incendios forestales

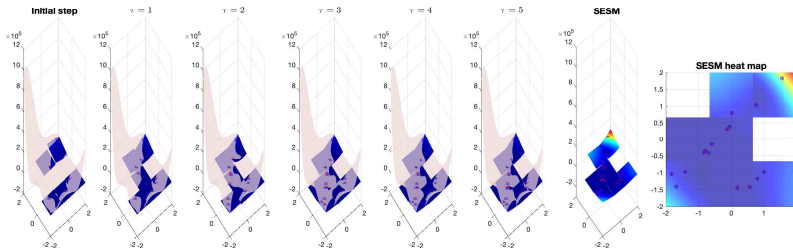
Modelo	SVR-Opt.		SVR-CM	
	MAD	RMSE	MAD	RMSE
STFWI	10.76	63.7	13.07	64.7
STM	11.51	64.0	13.13	64.7
FWI	11.76	64.0	12.86	64.7
M	11.01	63.7	12.71	64.7
All	10.59	63.5	No aplica	

Fig. Desviación absoluta promedio (MAD) y error cuadrático medio (RMSE) obtenidos al entrenar SVR-Optimizado y SVR-CM.

Conclusiones

Principal aporte

Diseño e implementación del modelo sustituto denominado SESM, que aproxima o ajusta funciones o aplicaciones a partir de pocas muestras, para ser utilizado dentro de un optimizador multiobjetivo.



Conclusiones

- El modo simultáneo presenta una mejor convergencia y consume un menor tiempo de ejecución que el secuencial.
- SESM, en comparación con RBF, presenta un mejor rendimiento al aproximar funciones o aplicaciones con cambios abruptos en sus máximos y mínimos locales y globales.
- Los experimentos demuestran que los hiperparámetros críticos de SESM son la cantidad de funciones base definidas en el diccionario y las iteraciones realizadas para ajustar los vectores dispersos.

Conclusiones

- La utilización de SESM dentro de un optimizador es relevante cuando los problemas están definidos a partir de funciones o aplicaciones de alto costo de evaluación.
- Un presupuesto muy bajo en la cantidad de muestras tomadas para ajustar los modelos sustitutos le da libertad al modelo y los resultados del mismo podrían llevar a soluciones inviables.
- La configuración óptima de todos los hiperparámetros de SESM depende de la función o aplicación a aproximar.

Preliminares
Antecedentes
Propuesta
Resultados
Conclusiones

Preguntas y comentarios...

Optimización

Escalarización	Métodos Clásicos
$F(x^*) \approx \min \sum_{k=1}^K w_k f_k(x)$ <ul style="list-style-type: none">■ Aproxima el frente de Pareto real■ Depende de las condiciones de escalarización establecidas	<ul style="list-style-type: none">■ Se derivan del desarrollo en series de Taylor■ Son exactos■ Dependen las derivadas de la función objetivo y de la convexidad del problema

Optimización

- Preliminares
- Antecedentes
- Propuesta
- Resultados
- Conclusiones

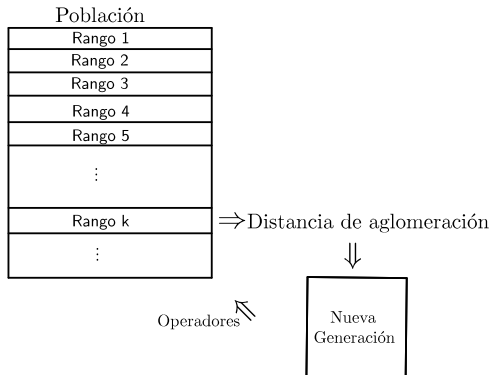
Algoritmos Evolutivos(Genéticos)

- Inspirados en la teoría evolutiva del origen de las especies
- No requieren condiciones de convexidad o diferenciación de las funciones objetivo

Optimización

NSGA-II (Fast Nondominated Sorting Genetic Algorithm)

Preliminares
Antecedentes
Propuesta
Resultados
Conclusiones



Optimización

Modelos sustitutos

- Modelos sustitutos, metamodelos o superficies de seudorespuesta reemplazan las funciones objetivo costosas por modelos de aproximación en el optimizador.
- Dado un conjunto de datos observados de f

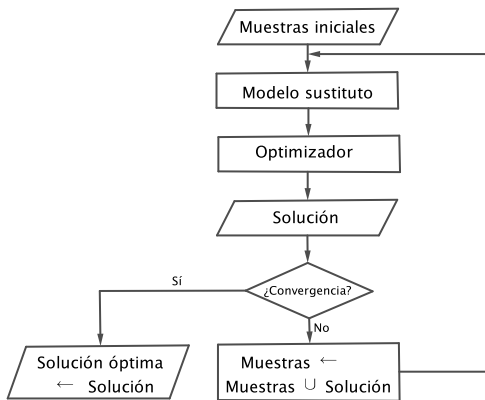
$$Y = f(X) = \{y^{(r)}\}_{r=1}^m = \{f(x^{(r)})\}_{r=1}^m$$

se aproxima f por medio de un modelo sustituto ψ

$$f(x) \approx \psi(x)$$

Optimización

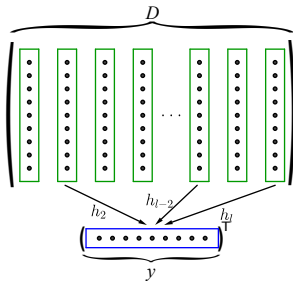
Modelos sustitutos



Codificación dispersa

Modela o reconstruye un vector y a partir de una combinación lineal dispersa h de los elementos de un marco (*frame*) llamado diccionario D

$$y \approx Dh$$



Codificación dispersa

Aprendizaje de diccionario (Dictionary Learning)

$$D^*, h^* = \arg \min_{D, h} \{ \|y - Dh\|_2^2 + \lambda \|h\|_1 \}$$



- ISTA: Iterative Shrinkage and Thresholding Algorithm

$$h^{(t)*} = \arg \min_{h^{(t)}} \|y^{(t)} - Dh^{(t)}\|_2^2 + \lambda \|h^{(t)}\|_1$$

- GD: Gradient Descent

$$D^* = \arg \min_D \sum_{t=1}^T \|y^{(t)} - Dh^{(t)}\|_2^2 = \arg \min_D \|Y - DH\|_F^2$$

Codificación dispersa

Aprendizaje de diccionario (Dictionary Learning)

$$D^*, h^* = \arg \min_{D, h} \{ \|y - Dh\|_2^2 + \lambda \|h\|_1 \}$$



- ISTA: Iterative Shrinkage and Thresholding Algorithm

$$h^{(t)*} = \arg \min_{h^{(t)}} \|y^{(t)} - Dh^{(t)}\|_2^2 + \lambda \|h^{(t)}\|_1$$

- GD: Gradient Descent

$$D^* = \arg \min_D \sum_{t=1}^T \|y^{(t)} - Dh^{(t)}\|_2^2 = \arg \min_D \|Y - DH\|_F^2$$

Propuesta

Diccionario paramétrico

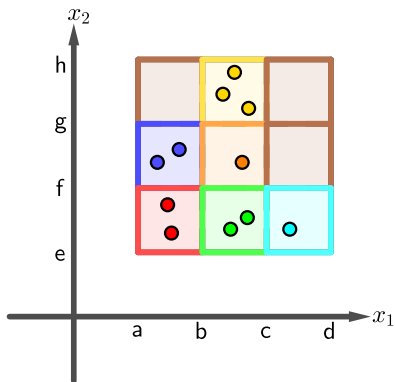
donde

$$\mu^{(j)} = \left(\mu_1^{(j)}, \mu_2^{(j)}, \mu_3^{(j)}, \dots, \mu_n^{(j)} \right)^T$$

$$\left(\Sigma^{(j)} \right)^{-1} = \begin{pmatrix} \rho_{1,1}^{(j)} & \rho_{1,2}^{(j)} & \rho_{1,3}^{(j)} & \cdots & \rho_{1,n-1}^{(j)} & \rho_{1,n}^{(j)} \\ \rho_{1,2}^{(j)} & \rho_{2,2}^{(j)} & \rho_{2,3}^{(j)} & \cdots & \rho_{2,n-1}^{(j)} & \rho_{2,n}^{(j)} \\ & & & \ddots & & \\ \rho_{1,n-1}^{(j)} & \rho_{2,n-1}^{(j)} & \rho_{3,n-1}^{(j)} & \cdots & \rho_{n-1,n-1}^{(j)} & \rho_{n-1,n}^{(j)} \\ \rho_{1,n}^{(j)} & \rho_{2,n}^{(j)} & \rho_{3,n}^{(j)} & \cdots & \rho_{n-1,n}^{(j)} & \rho_{n,n}^{(j)} \end{pmatrix}$$

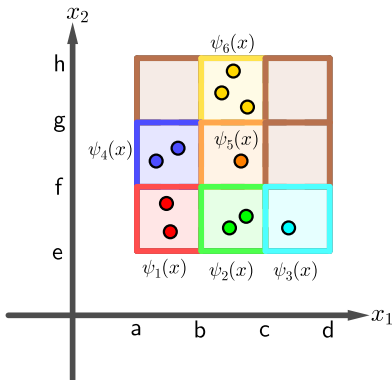
Propuesta

SESM: Partición



Propuesta

SESM: Partición



- 1 $\psi_1(x) = D(x) \cdot h^{(1)}$
- 2 $\psi_2(x) = D(x) \cdot h^{(2)}$
- 3 $\psi_3(x) = D(x) \cdot h^{(3)}$
- 4 $\psi_4(x) = D(x) \cdot h^{(4)}$
- 5 $\psi_5(x) = D(x) \cdot h^{(5)}$
- 6 $\psi_6(x) = D(x) \cdot h^{(6)}$

Propuesta

SESM: Partición

$$1 \quad \psi_1(x) \leftarrow \underbrace{(\phi_1(X^{(1)}) \quad \phi_2(X^{(1)}) \quad \phi_3(X^{(1)}) \quad \phi_4(X^{(1)}) \quad \phi_5(X^{(1)}))}_{D^{(1)}}$$

$$2 \quad \psi_2(x) \leftarrow \underbrace{(\phi_1(X^{(2)}) \quad \phi_2(X^{(2)}) \quad \phi_3(X^{(2)}) \quad \phi_4(X^{(2)}) \quad \phi_5(X^{(2)}))}_{D^{(2)}}$$

$$3 \quad \psi_3(x) \leftarrow \underbrace{(\phi_1(X^{(3)}) \quad \phi_2(X^{(3)}) \quad \phi_3(X^{(3)}) \quad \phi_4(X^{(3)}) \quad \phi_5(X^{(3)}))}_{D^{(3)}}$$

$$4 \quad \psi_4(x) \leftarrow \underbrace{(\phi_1(X^{(4)}) \quad \phi_2(X^{(4)}) \quad \phi_3(X^{(4)}) \quad \phi_4(X^{(4)}) \quad \phi_5(X^{(4)}))}_{D^{(4)}}$$

$$5 \quad \psi_5(x) \leftarrow \underbrace{(\phi_1(X^{(5)}) \quad \phi_2(X^{(5)}) \quad \phi_3(X^{(5)}) \quad \phi_4(X^{(5)}) \quad \phi_5(X^{(5)}))}_{D^{(5)}}$$

$$6 \quad \psi_6(x) \leftarrow \underbrace{(\phi_1(X^{(6)}) \quad \phi_2(X^{(6)}) \quad \phi_3(X^{(6)}) \quad \phi_4(X^{(6)}) \quad \phi_5(X^{(6)}))}_{D^{(6)}}$$

Propuesta

SESM: Partición

$$1 \quad \psi_1(x) \leftarrow \underbrace{(\phi_1(X^{(1)}) \quad \phi_2(X^{(1)}) \quad \phi_3(X^{(1)}) \quad \phi_4(X^{(1)}) \quad \phi_5(X^{(1)}))}_{D^{(1)}}$$

$$2 \quad \psi_2(x) \leftarrow \underbrace{(\phi_1(X^{(2)}) \quad \phi_2(X^{(2)}) \quad \phi_3(X^{(2)}) \quad \phi_4(X^{(2)}) \quad \phi_5(X^{(2)}))}_{D^{(2)}}$$

$$3 \quad \psi_3(x) \leftarrow \underbrace{(\phi_1(X^{(3)}) \quad \phi_2(X^{(3)}) \quad \phi_3(X^{(3)}) \quad \phi_4(X^{(3)}) \quad \phi_5(X^{(3)}))}_{D^{(3)}}$$

$$4 \quad \psi_4(x) \leftarrow \underbrace{(\phi_1(X^{(4)}) \quad \phi_2(X^{(4)}) \quad \phi_3(X^{(4)}) \quad \phi_4(X^{(4)}) \quad \phi_5(X^{(4)}))}_{D^{(4)}}$$

$$5 \quad \psi_5(x) \leftarrow \underbrace{(\phi_1(X^{(5)}) \quad \phi_2(X^{(5)}) \quad \phi_3(X^{(5)}) \quad \phi_4(X^{(5)}) \quad \phi_5(X^{(5)}))}_{D^{(5)}}$$

$$6 \quad \psi_6(x) \leftarrow \underbrace{(\phi_1(X^{(6)}) \quad \phi_2(X^{(6)}) \quad \phi_3(X^{(6)}) \quad \phi_4(X^{(6)}) \quad \phi_5(X^{(6)}))}_{D^{(6)}}$$

Propuesta

SESM: Partición

$$1 \quad \psi_1(x) \leftarrow \underbrace{(\phi_1(X^{(1)}) \quad \phi_2(X^{(1)}) \quad \phi_3(X^{(1)}) \quad \phi_4(X^{(1)}) \quad \phi_5(X^{(1)}))}_{D^{(1)}}$$

$$2 \quad \psi_2(x) \leftarrow \underbrace{(\phi_1(X^{(2)}) \quad \phi_2(X^{(2)}) \quad \phi_3(X^{(2)}) \quad \phi_4(X^{(2)}) \quad \phi_5(X^{(2)}))}_{D^{(2)}}$$

$$3 \quad \psi_3(x) \leftarrow \underbrace{(\phi_1(X^{(3)}) \quad \phi_2(X^{(3)}) \quad \phi_3(X^{(3)}) \quad \phi_4(X^{(3)}) \quad \phi_5(X^{(3)}))}_{D^{(3)}}$$

$$4 \quad \psi_4(x) \leftarrow \underbrace{(\phi_1(X^{(4)}) \quad \phi_2(X^{(4)}) \quad \phi_3(X^{(4)}) \quad \phi_4(X^{(4)}) \quad \phi_5(X^{(4)}))}_{D^{(4)}}$$

$$5 \quad \psi_5(x) \leftarrow \underbrace{(\phi_1(X^{(5)}) \quad \phi_2(X^{(5)}) \quad \phi_3(X^{(5)}) \quad \phi_4(X^{(5)}) \quad \phi_5(X^{(5)}))}_{D^{(5)}}$$

$$6 \quad \psi_6(x) \leftarrow \underbrace{(\phi_1(X^{(6)}) \quad \phi_2(X^{(6)}) \quad \phi_3(X^{(6)}) \quad \phi_4(X^{(6)}) \quad \phi_5(X^{(6)}))}_{D^{(6)}}$$

Propuesta

SESM: Partición

$$1 \quad \psi_1(x) \leftarrow \underbrace{(\phi_1(X^{(1)}) \quad \phi_2(X^{(1)}) \quad \phi_3(X^{(1)}) \quad \phi_4(X^{(1)}) \quad \phi_5(X^{(1)}))}_{D^{(1)}}$$

$$2 \quad \psi_2(x) \leftarrow \underbrace{(\phi_1(X^{(2)}) \quad \phi_2(X^{(2)}) \quad \phi_3(X^{(2)}) \quad \phi_4(X^{(2)}) \quad \phi_5(X^{(2)}))}_{D^{(2)}}$$

$$3 \quad \psi_3(x) \leftarrow \underbrace{(\phi_1(X^{(3)}) \quad \phi_2(X^{(3)}) \quad \phi_3(X^{(3)}) \quad \phi_4(X^{(3)}) \quad \phi_5(X^{(3)}))}_{D^{(3)}}$$

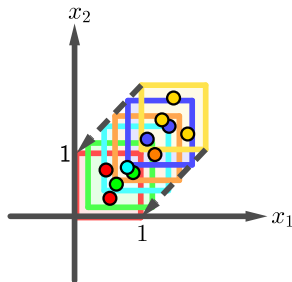
$$4 \quad \psi_4(x) \leftarrow \underbrace{(\phi_1(X^{(4)}) \quad \phi_2(X^{(4)}) \quad \phi_3(X^{(4)}) \quad \phi_4(X^{(4)}) \quad \phi_5(X^{(4)}))}_{D^{(4)}}$$

$$5 \quad \psi_5(x) \leftarrow \underbrace{(\phi_1(X^{(5)}) \quad \phi_2(X^{(5)}) \quad \phi_3(X^{(5)}) \quad \phi_4(X^{(5)}) \quad \phi_5(X^{(5)}))}_{D^{(5)}}$$

$$6 \quad \psi_6(x) \leftarrow \underbrace{(\phi_1(X^{(6)}) \quad \phi_2(X^{(6)}) \quad \phi_3(X^{(6)}) \quad \phi_4(X^{(6)}) \quad \phi_5(X^{(6)}))}_{D^{(6)}}$$

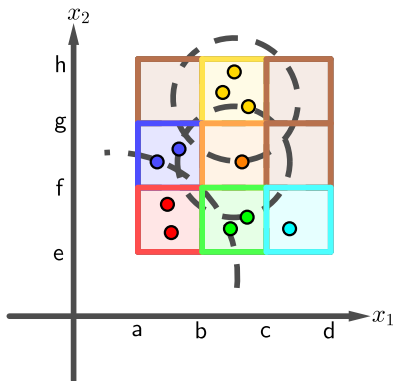
Propuesta

SESM: Normalización



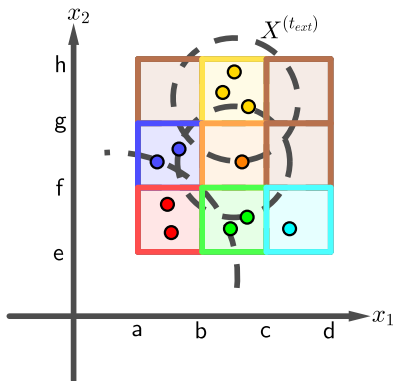
Propuesta

SESM: Continuidad



Propuesta

SESM: Continuidad



Resultados

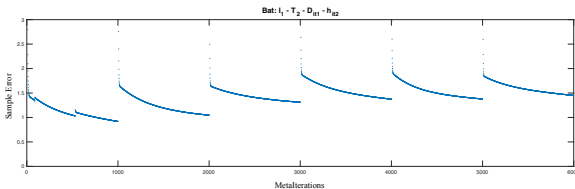
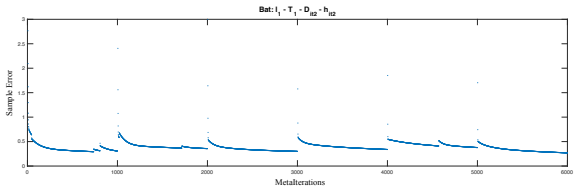
- Preliminares
- Antecedentes
- Propuesta
- Resultados
- Conclusiones

Variabilidad: Comportamiento de SESM ante diferentes configuraciones

- Funciones definidas en \mathbb{R}^2
- Hiperparámetros:
 - Modo del algoritmo
 - Cantidad de palabras en el diccionario l
 - Partición del espacio T
 - Iteraciones para ajustar al diccionario D_{iter}
 - Iteraciones para ajustar al vector disperso h_{iter}

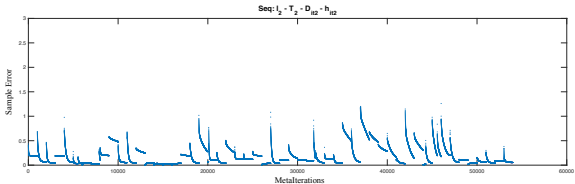
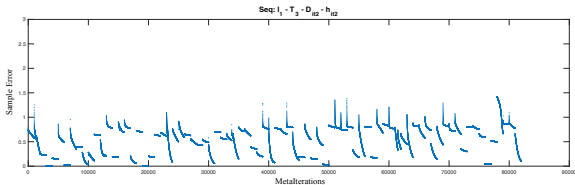
Resultados

Variabilidad: Modo simultáneo

 $I \in \{50, 150\}$; $T \in \{1, 2, 3\}$; $D_{iter} \in \{1, 15\%(\text{Param.Cod.})\}$; $h_{iter} \in \{1, 15\%I\}$ 

Resultados

Variabilidad: Modo secuencial

$$l \in \{50, 150\}; T \in \{1, 2, 3\}; D_{iter} \in \{1, 15\%(\text{Param.Cod.})\}; h_{iter} \in \{1, 15\%l\}$$


Resultados

Eficiencia

- Preliminares
- Antecedentes
- Propuesta
- Resultados
- Conclusiones

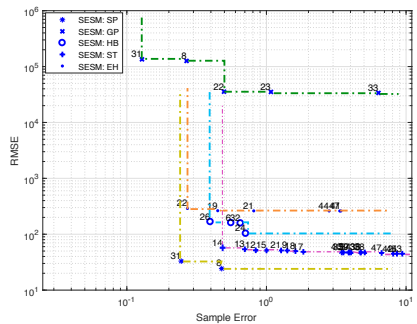


Fig. Frentes de Pareto, dado por los errores locales y globales obtenidos con SESM