



# Una metaheurística paralela para grandes problemas de optimización dinámica entera mixta, con aplicaciones en la biología computacional

David R Penas

*Departamento de Estadística, Análisis matemático y Optimización, y Instituto de Matemáticas (IMAT) Universidade de Santiago de Compostela (USC), España david.rodriguez.penas@usc.es*

David Henriques

*Instituto de Investigaciones Marinas (IIM) Consejo Superior de Investigaciones Científicas (CSIC) Vigo, España*

Patricia González

*Grupo de Arquitectura de Computadores (GAC) Universidade da Coruña (UDC) Coruña, España*

Ramón Doallo

*Grupo de Arquitectura de Computadores (GAC) Universidade da Coruña (UDC) Coruña, España*

Julio Saez-Rodriguez

*Joint Research Centre for Computational Biomedicine (JCR-Combine) RWTH Aachen University Aquisgrán, Alemania*

Julio R Banga

*Instituto de Investigaciones Marinas (IIM) Consejo Superior de Investigaciones Científicas (CSIC) Vigo, España*

**Abstract**—Esto es un resumen de nuestro artículo publicado en PLoS ONE [1] para su presentación en la Multiconferencia CAEPIA'18 Key Works.

**Index Terms**—Optimización MIDO-MINLP, programación paralela, metaheurística cooperativa, Scatter Search.

## I. INTRODUCCIÓN

Muchos problemas clave dentro de la bioinformática, la biología computacional o la biología de sistemas [2], se pueden resolver a través de formulaciones basadas en la optimización matemática. Un ejemplo es el desarrollo de modelos matemáticos dinámicos, donde se mapean comportamientos que varían en el tiempo de determinados procesos biológicos, como por ejemplo el de rutas metabólicas, o la regulación de determinados genes.

En este trabajo nos hemos centrado en los conocidos en bioinformática como problemas de ingeniería inversa [3], los cuales consisten en reconstruir la estructura de una red celular a partir de sus datos experimentales. Este proceso puede ser formulado como un problema de optimización dinámica entera mixta (MIDO), debido a las variables dependientes en el tiempo y al uso de logic-ODEs [4] para mapear ciertos comportamientos en el modelo, y donde además parte de las variables de decisión son discretas (binarias o enteras). Adicionalmente, este tipo de optimización también pertenece a la programación no lineal entera mixta (MINLP), dando lugar a problemas extremadamente difíciles de resolver debido a que presentan no linealidad, no convexidad, mal acondicionamiento y multimodalidad.

Las metaheurísticas son una buena alternativa a tener en cuenta en este contexto, ya que generalmente alcanzan una

región cercana a la solución global en un tiempo de ejecución aceptable. Sin embargo, cuando estos problemas tienen un tamaño o complejidad considerable, el costo computacional puede llegar a ser muy alto [5], [6].

Por ello, en este trabajo se ha presentado saCeSS2 [7]: una metaheurística cooperativa y auto-adaptativa con capacidad para resolver problemas MIDO-MINLP de manera satisfactoria en un tiempo de cálculo razonable.

## II. MÉTODOS

Una de las contribuciones de este trabajo es la actualización de saCeSS [8], la cual es una herramienta distribuida, cooperativa y auto-adaptativa del método de búsqueda dispersa, en concreto del enhanced Scatter Search (eSS) [9], [10]. Así, saCeSS es un optimizador global orientado a trabajar con problemas NLP que implementa las siguientes funcionalidades:

- *Paralelización a grano fino combinando el modelo maestro-esclavo con un modelo en islas:* cada esclavo/isla ejecuta de manera aislada una búsqueda con eSS, y cuando se encuentra una solución prometedora, se envía al maestro, el cual es el encargado de gestionar las comunicaciones entre las diferentes islas, evitando así un exceso de compartición de resultados que podría degradar la diversidad del método.
- *Cooperación de soluciones basado en la calidad de la solución:* el intercambio de información entre islas solo se realiza cuando una isla encuentra una solución lo suficientemente buena.
- *Comunicaciones asíncronas:* el protocolo de comunicaciones diseñado es asíncrono, para evitar retardos y

esperas, haciendo que la búsqueda de soluciones de cada isla nunca se pare.

- *Esquema autoadaptativo*: se ha implementado un mecanismo de auto-configuración de los parámetros de la metaheurística. Cada isla comenzará el proceso de optimización con unas opciones de configuración diferentes. A medida que la búsqueda avance, el maestro permitirá que los esclavos que hayan aportado menos soluciones a la búsqueda global adquieran la configuración de aquellos que hayan sido más exitosos.

Por lo tanto, la principal aportación de este trabajo es la presentación de un nuevo método, saCeSS2, que resulta de las modificaciones del método saCeSS original con el principal objetivo de resolver problemas MIDO-MINLP. Esta extensión de funcionalidades ha seguido tres direcciones principales:

- 1) *Inclusión de un solucionador local especializado en la resolución de problemas de optimización entera mixta*: se ha combinado el eSS con un método de programación cuadrática, llamado Mixed-Integer Sequential Quadratic Programming (MISQP) [11], en cada isla.
- 2) *Modificación del mecanismo de auto-adaptación con el objetivo de evitar estancamiento prematuro de la convergencia*: en este tipo de problemas de optimización, encontrar nuevas soluciones candidatas a ser compartidas no es fácil, por lo tanto se han relajado las condiciones de adaptación del método.
- 3) *Implementación de nuevos mecanismo para mantener la diversidad durante el proceso de cooperación*: se ha observado que en este tipo de problemas hay que ser muy cautos durante la compartición de soluciones, ya que una cooperación agresiva daña muy rápidamente la diversidad de la isla. Por eso, cuando se realice un cambio en los parámetros de configuración de una isla que no ha contribuido lo suficiente, se produce un reinicio aleatorio de la mayoría de soluciones de su población.

Otra de las contribuciones de este trabajo ha sido la extensión y generalización de la formulación considerada por Henriques et al [4] al adoptar un enfoque genérico del problema del control óptimo entero mixto, pero sin utilizar relajaciones o transformaciones del problema original durante su resolución.

### III. RESULTADOS

Se ha probado el método propuesto en tres casos de estudio de ingeniería inversa muy complejos relacionados con la señalización celular: (1) un problema que considera una vía de señalización sintética, con 84 variables de decisión continuas y 34 binarias; (2) un modelo dinámico de señalización en cáncer de hígado que utiliza datos de alto rendimiento, con 135 variables de decisión continuas y 109 binarias; y (3) un problema extremadamente difícil relacionado con el cáncer de mama, con 690 variables de decisión continuas y 138 binarias.

Se reportaron resultados computacionales obtenidos en diferentes infraestructuras, incluyendo un clúster local, un

gran superordenador y una plataforma de Cloud pública. Curiosamente, los resultados muestran cómo la cooperación de búsquedas paralelas individuales modifica las propiedades sistémicas del algoritmo secuencial, logrando aceleraciones superlineales en comparación con una búsqueda individual (por ejemplo, aceleraciones de 15 con 10 núcleos) y mejorando significativamente (por encima del 60%) el rendimiento con respecto a un esquema paralelo no cooperativo. La escalabilidad del método también es buena (las pruebas se realizaron usando hasta 300 núcleos). Adicionalmente, se ha obtenido que la utilización de saCeSS2 en una infraestructura pública de Cloud tiene un buen rendimiento en configuraciones que usan 10 núcleos, pero el rendimiento se degrada considerablemente cuando se trata de utilizar más procesadores.

### IV. CONCLUSIONES

Los resultados mostrados en este trabajo demuestran que saCeSS2 puede ser útil en los problemas de ingeniería inversa en modelos dinámicos de complejas rutas biológicas. Además, los éxitos obtenidos mediante esta herramienta pueden ser fácilmente exportados a otros casos de estudios que también generen problemas de optimización MIDO-MINLP a gran escala, como por ejemplo la ingeniería metabólica, la biología sintética, o la programación de fármacos.

Por otro lado, desde el punto de vista computacional, también se han demostrado que las estrategias de virtualización y computación en Cloud, pueden ser una buena alternativa, aunque presentan unas bien conocidas limitaciones debido principalmente a sobrecargas en la compartición de recursos.

### REFERENCES

- [1] D.R. Penas, D. Henriques, P. González, R. Doallo, J. Saez-Rodriguez and J.R. Banga, "A parallel metaheuristic for large mixed-integer dynamic optimization problems, with applications in computational biology" *PLoS ONE*, 12(8): e0182186, 2017.
- [2] J.R. Banga, "Optimization in computational systems biology" *BMC Systems Biology*, 2 (1), 47, 2008.
- [3] A.F. Villaverde and J.R. Banga, "Reverse engineering and identification in systems biology: strategies, perspectives and challenges" *Journal of The Royal Society Interface*, 11(91), 20130505, 2014.
- [4] D. Henriques, A.F. Villaverde, M. Rocha, J. Saez-Rodriguez, J.R. Banga "Data-driven reverse engineering of signaling pathways using ensembles of dynamic models" *PLoS Computational Biology*, 13(2), e1005379, 2017.
- [5] P. Mendes and D. Kell "Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation", *Bioinformatics*, 14 (10), 869–883, 1998.
- [6] C.G. Moles, P. Mendes and J.R. Banga, "Parameter estimation in biochemical pathways: a comparison of global optimization methods", *Genome Research*, 13 (11), 2467–2474, 2003.
- [7] saCeSS2 - global mixed-integer optimization library version 2017A, <https://zenodo.org/record/290219>
- [8] D.R. Penas, P. González, J.A. Egea, R. Doallo and J.R. Banga "Parameter estimation in large-scale systems biology models: a parallel and self-adaptive cooperative strategy", *BMC Bioinformatics*, 18:52, 2017.
- [9] J.A. Egea, R. Martí, J.R. Banga, "An evolutionary method for complex-process optimization", *Computers & Operations Research*, 37(2), 315–324, 2010.
- [10] J.A. Egea, E. Balsa-Canto, M.S. G. García, J.R. Banga, "Dynamic optimization of nonlinear processes with an enhanced scatter search method", *Industrial & Engineering Chemistry Research*, 48(9), 4388–4401, 2009.
- [11] O. Exler, K. Schittkowski, "A trust region sqp algorithm for mixed-integer nonlinear programming", *Optimization Letters*, 1(3), 269–280, 2007.