A Regularized Interior-Point Method for Constrained Nonlinear Least Squares

XII Brazilian Workshop on Continuous Optimization

Abel Soares Siqueira
Federal University of Paraná - Curitiba/PR - Brazil

Dominique Orban
GERAD/Polytechnique Montréal - Montréal - Canada

July 23, 2018

Problem

$$
\begin{aligned}
\text{minimize} \quad & f(x) = \tfrac{1}{2}\|F(x)\|^2 \\
\text{subject to} \quad & c(x) = 0, \\
& \ell \le x \le u,
\end{aligned}
\qquad \text{(CNLS)}
$$

where $F : \mathbb{R}^n \to \mathbb{R}^{n_E}$ and $c : \mathbb{R}^n \to \mathbb{R}^m$ are $\mathcal{C}^2$.

Problem

$$\begin{aligned} \text{minimize} \quad & f(x) = \frac{1}{2}\|F(x)\|^2 \\ \text{subject to} \quad & c(x) = 0, \\ & x \geq 0, \end{aligned} \qquad \text{(CNLS)}$$

where $F : \mathbb{R}^n \to \mathbb{R}^{n_E}$ and $c : \mathbb{R}^n \to \mathbb{R}^m$ are $\mathcal{C}^2$.

Friedlander and Orban [5] primal-dual exact regularization

Quadratic programming primal

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}x^TQx + c^Tx \\ \text{subject to} & Ax = b, \qquad x \geq 0. \end{array} \tag{QP}$$

Dual of (QP)

$$\begin{array}{ll} \text{maximize} & b^Ty - \frac{1}{2}x^TQx \\ \text{subject to} & -Qx + A^Ty + z = c, \qquad z \geq 0. \end{array} \tag{QD}$$

Friedlander and Orban [5] primal-dual exact regularization

Regularized quadratic programming primal

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}x^TQx + c^Tx + \frac{1}{2}\rho\|x - x_k\|^2 + \frac{1}{2}\delta\|r + y_k\|^2 \\ \text{subject to} & Ax + \delta r = b, \qquad x \geq 0. \end{array} \tag{RP}$$

Regularized dual of (QP) - similar to dual of (RP)

$$\begin{array}{ll} \text{maximize} & b^Ty - \frac{1}{2}x^TQx - \frac{1}{2}\delta\|y - y_k\|^2 - \frac{1}{2}\rho\|s + x_k\|^2 \\ \text{subject to} & -Qx + A^Ty + z - \rho s = c, \qquad z \geq 0. \end{array} \tag{RD}$$

Arreckx and Orban [2]

$$\begin{array}{ll} \text{minimize} & f(x) + \frac{1}{2}\rho\|x - x_k\|^2 + \frac{1}{2}\delta\|u + y_k\|^2 \\ \text{subject to} & c(x) + \delta u = 0. \end{array} \tag{1}$$

Dehghani et al. [3]

$$\begin{array}{ll} \text{minimize} & c^T x + \frac{1}{2}\|Cx - d\|^2 + \frac{1}{2}\rho\|x - x_k\|^2 + \frac{1}{2}\delta\|u + y_k\|^2 \\ \text{subject to} & Ax + \delta u = b, \qquad x \geq 0. \end{array} \tag{2}$$

Regularization of (CNLS)

$$
\begin{aligned}
\text{minimize} \quad & \tfrac{1}{2}\|F(x)\|^2 + \tfrac{1}{2}\rho\|x - x_k\|^2 + \tfrac{1}{2}\delta\|u + y_k\|^2 \\
\text{subject to} \quad & c(x) + \delta u = 0, \\
& x \geq 0.
\end{aligned} \tag{3}
$$

Regularization of (CNLS)

$$
\begin{aligned}
\text{minimize} \quad & \frac{1}{2}\|r\|^2 + \frac{1}{2}\rho\|x - x_k\|^2 + \frac{1}{2}\delta\|u + y_k\|^2 \\
\text{subject to} \quad & F(x) - r = 0 \\
& c(x) + \delta u = 0, \\
& x \geq 0.
\end{aligned}
\tag{3}
$$

KKT

$$
\begin{aligned}
\rho(x - x_k) - A(x)^T w^r - B(x)^T y - z &= 0 \\
r + w^r &= 0 \\
\delta(u + y_k) - \delta y &= 0 \\
F(x) - r &= 0 \\
c(x) + \delta u &= 0 \\
Xz &= 0 \\
(x, z) &\geq 0
\end{aligned}
$$

$$A(x) = \nabla F(x), \qquad B(x) = \nabla c(x).$$

KKT

$$
\begin{aligned}
\rho(x - x_k) + A(x)^T r - B(x)^T y - z &= 0 \\
F(x) - r &= 0 \\
c(x) + \delta(y - y_k) &= 0 \\
Xz &= 0 \\
(x, z) &\geq 0
\end{aligned}
$$

$$
A(x) = \nabla F(x), \qquad B(x) = \nabla c(x).
$$

KKT

$$G_k(x,r,y,z) = \begin{bmatrix} \rho(x - x_k) + A(x)^T r - B(x)^T y - z \\ F(x) - r \\ c(x) + \delta(y - y_k) \\ Xz \end{bmatrix}.$$

KKT

$$G_k(\underbrace{x, r, y, z}_{w}) = \begin{bmatrix} \rho(x - x_k) + A(x)^T r - B(x)^T y - z \\ F(x) - r \\ c(x) + \delta(y - y_k) \\ Xz \end{bmatrix}.$$

$$w_k = (x_k, r_k, \lambda_k, z_k) \qquad \Delta w = (\Delta x_k, \Delta r_k, \Delta\lambda_k, \Delta z_k)$$

KKT

$$G_k(w_k) = \begin{bmatrix} A_k^T r_k - B_k^T y_k - z_k \\ F(x_k) - r_k \\ c(x_k) \\ X_k z_k \end{bmatrix}.$$

KKT

$$J_{G_k}(w_k) = \begin{bmatrix} H_k & A_k^T & -B_k^T & -I \\ A_k & -I & 0 & 0 \\ B_k & 0 & \delta I & 0 \\ Z_k & 0 & 0 & X_k \end{bmatrix}$$

$$H_k = \rho I + \sum_{i=1}^{n_E} \nabla^2 F_i(x_k)(r_k)_i - \sum_{i=1}^{m} \nabla^2 c_i(x_k)(y_k)_i.$$

Newton interior point step

$$J_{G_k}(w_k)\Delta w = -G_k(w_k) + \mu_k \tilde{e}$$

$$\tilde{e}^T = (0, 0, 0, e^T)$$

$$\begin{bmatrix} H_k & A_k^T & -B_k^T & -I \\ A_k & -I & 0 & 0 \\ B_k & 0 & \delta I & 0 \\ Z_k & 0 & 0 & X_k \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta r \\ \Delta y \\ \Delta z \end{bmatrix} = \begin{bmatrix} z_k - A_k^T r_k + B_k^T y_k \\ r_k - F(x_k) \\ -c(x_k) \\ \mu_k e - X_k z_k \end{bmatrix}$$

Symmetric Quasi-Definite systems

Symmetric Quasi-Definite: K is SQD if there is some permutation matrix P such that

$$P^T K P = \begin{bmatrix} M & A^T \\ A & -N \end{bmatrix},$$

where M and N are symmetric positive definite.

Symmetric Quasi-Definite systems

Theorem (Vanderbei [10])

An SQD matrix is strongly factorizable, that is, if K is SQD, then for any permutation matrix P, there are matrices L and D such that

$$P^T K P = LDL^T,$$

where L is lower triangular with unit diagonal and D is diagonal.

Symmetric Quasi-Definite systems

$$\begin{bmatrix} H_k & A_k^T & B_k^T & -Z_k^{1/2} \\ A_k & -I & 0 & 0 \\ B_k & 0 & -\delta I & 0 \\ -Z_k^{1/2} & 0 & 0 & -X_k \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta r \\ -\Delta y \\ Z_k^{-1/2}\Delta z \end{bmatrix} = \begin{bmatrix} z_k - A_k^T r_k + B_k^T y_k \\ r_k - F(x_k) \\ -c(x_k) \\ Z_k^{1/2} x_k - \mu_k Z_k^{-1/2} e \end{bmatrix}$$

Symmetric Quasi-Definite systems

$$\begin{bmatrix} H_k + X_k^{-1} Z_k & A_k^T & B_k^T \\ A_k & -I & 0 \\ B_k & 0 & -\delta I \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta r \\ -\Delta y \end{bmatrix} = \begin{bmatrix} \mu_k X_k^{-1} e - A_k^T r_k + B_k^T y_k \\ r_k - F(x_k) \\ -c(x_k) \end{bmatrix}$$

Framework of Armand et al. [1]
k-th iteration

1: Choose $\mu_k^+ > 0$ and $\tau_k \in (0,1)$.

2: Compute Δw_k, solution of

$$J_{G_k}(w_k)\Delta w_k + G_k(w_k) - \mu_k^+ \tilde{e} = 0.$$

3: Compute α_k as the largest $\alpha \in (0,1]$ such that

$$(x_k + \alpha \Delta x_k, z_k + \alpha \Delta z_k) \geq (1 - \tau_k)(x_k, z_k).$$

4: Choose $a_k = (a_k^x, a_k^r, a_k^\lambda, a_k^z) \in [\alpha_k, 1]^N$ such that $\begin{cases} x_k + a_k^x . * \Delta x_k > 0 \\ z_k + a_k^z . * \Delta z_k > 0. \end{cases}$

5: Set $\hat{\mu}_k = \mu_k + \alpha_k(\mu_k^+ - \mu_k)$ and $\hat{w}_k = w_k + a_k . * \Delta w_k$.

Choose μ_{k+1} between μ_k^+ and $\hat{\mu}_k$, and choose $\epsilon_k > 0$.

if $\|G_{k+1}(\hat{w}_k) - \mu_{k+1}\tilde{e}\| \leq \theta\|G_k(w_k) - \mu_k\tilde{e}\| + \epsilon_k$ **then**

  $w_{k+1} = \hat{w}_k$,

else

  Perform inner iterations with barrier parameter μ_{k+1} to identify w_{k+1} such that $(x_{k+1}, z_{k+1}) > 0$ and $\|G_{k+1}(w_{k+1}) - \mu_{k+1}\tilde{e}\| \leq \theta\|G_k(w_k) - \mu_k\tilde{e}\| + \epsilon_k$.

end if

Global convergence

Assumptions

A1 The sequences $\{H_k\}$, $\{A_k\}$ and $\{B_k\}$ are bounded.

A2 $\delta_k = \Omega(\mu_k)$.

A3 The matrices $H_k + X_k^{-1} Z_k + \rho_k I + \frac{1}{\delta_k} A_k^T A_k + B_k^T B_k$ are uniformly positive definite for $k \in \mathbb{N}$.

A4 The sequences $\{\mu_k^+\}$ and $\{\mu_k\}$ satisfy $\limsup_{k \to \infty} \frac{\mu_k^+}{\mu_k} < 1$.

A5 The inner iteration are globally convergent, i.e., we can always find w_{k+1}.

Global convergence

Theorem (3.1 of Armand et al. [1], adjusted to our problem)

Assume A1-A5, that $\{\tau_k\}$ *is bounded away from zero, and* $\epsilon_k \to 0$. *Then the algorithm generates a sequence* $\{w_k\}$ *such that* $\{\mu_k\}$ *and* $\{G_k(w_k)\}$ *converge to zero.*

Implementation

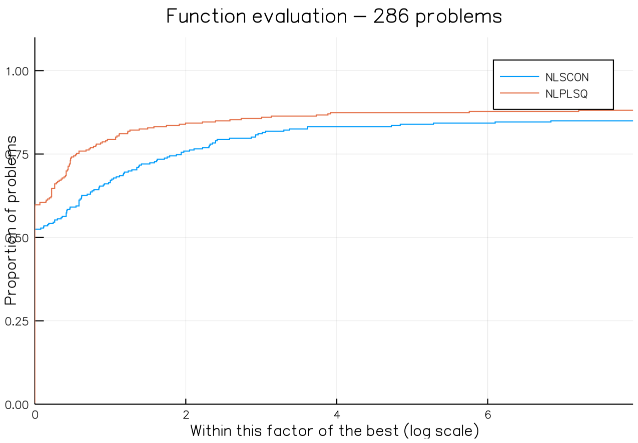
- Implemented in Julia with the JuliaSmoothOptimizers [7] tools;
- Regularization update following Wächter and Biegler [11];
- System solved with LDL factorization;
- Approximation of the Hessian using Dennis et al. [4];
- Inner iterations consist of line search on Merit Function

$$\psi(x, r, \lambda; \eta) = \frac{1}{2}\|r\|^2 - \mu \sum_{i=1}^{n} \ln x_i + \frac{\rho}{2}\|x - x_k\|^2 + \frac{\delta}{2}\|\lambda\|^2 + \\ + \eta\Big[\|c(x) + \delta(\lambda - \lambda_k)\|_1 + \|F(x) - r\|_1\Big].$$

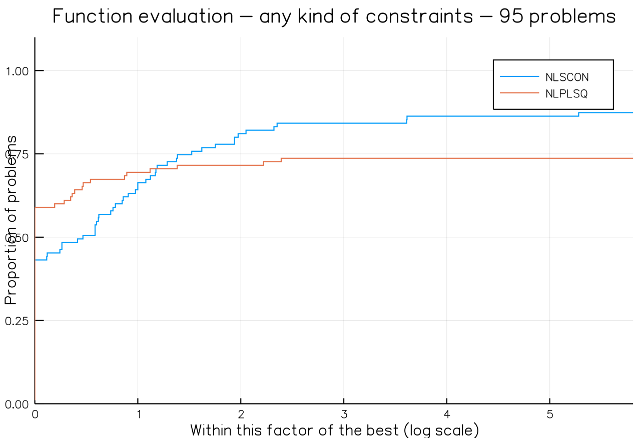
Comparison

- Preliminary results;
- Comparison against NLPLSQ [8, 9], which uses similar approach;
- Compared with 286 problems from NLSProblems (part of [7]) and CUTEst [6] NLS problems;
- 191 problems are unconstrained;
- 17 problems are only bounded;
- 43 problems have only equality constraints;
- 35 problems are more general;

Comparison


Function evaluation – 286 problems
NLSCON
NLPLSQ
1.00
0.75
0.50
0.25
0.00
Proportion of problems
0
2
4
6
Within this factor of the best (log scale)

Comparison


Function evaluation — any kind of constraints — 95 problems
NLSCON
NLPLSQ
Proportion of problems
1.00
0.75
0.50
0.25
0.00
0
1
2
3
4
5
Within this factor of the best (log scale)

Future work

- Factorization-free implementation;
- Extension for $f(x) = g(x) + \frac{1}{2}\|F(x)\|^2$;
- Large scale application.

References

[1] P. Armand, J. Benoist, and D. Orban. From global to local convergence of interior methods for nonlinear optimization. *Optimization Methods and Software*, 28(5): 1051–1080, 2013. doi: 10.1080/10556788.2012.668905.

[2] S. Arreckx and D. Orban. A regularized factorization-free method for equality-constrained optimization. *SIAM Journal on Optimization*, pages –, 2018.

[3] M. Dehghani, A. Lambe, and D. Orban. A regularized interior-point method for constrained linear least squares. Technical Report G-2018-07, GERAD, HEC Montréal, Canada, 2018.

[4] J. E. Dennis, Jr., D. M. Gay, and R. E. Walsh. An adaptive nonlinear least-squares algorithm. *ACM Transactions on Mathematical Software*, 7(3):348–368, 1981. doi: 10.1145/355958.355965.

[5] M. P. Friedlander and D. Orban. A primal-dual regularized interior-point method for convex quadratic programs. *Mathematical Programming Computation*, 4(1):71–107, 2012. doi: 10.1007/s12532-012-0035-2.

[6] N. I. Gould, D. Orban, and P. L. Toint. CUTEst: A constrained and unconstrained testing environment with safe threads for mathematical optimization. *Comput. Optim. Appl.*, 60 (3):545–557, 2015. doi: 10.1007/s10589-014-9687-3.

[7] D. Orban and A. S. Siqueira. JuliaSmoothOptimizers. URL `https://JuliaSmoothOptimizers.github.io`.

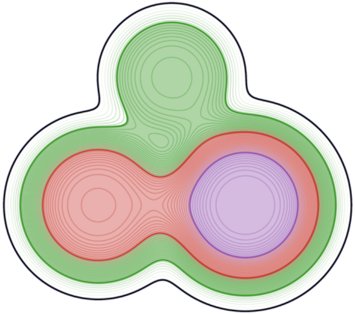
[8] K. Schittkowski. *Solving Constrained Nonlinear Least Squares Problems by a General Purpose SQP-Method*, pages 295–309. Birkhäuser Basel, Basel, 1988. ISBN 978-3-0348-9297-1. doi: 10.1007/978-3-0348-9297-1_19.

[9] K. Schittkowski. NLPLSQ: A fortran implementation of an SQP-Gauss-Newton algorithm for least squares optimization. Technical report, Department of Computer Science, University of Bayreuth, 2007.

[10] R. J. Vanderbei. Symmetric quasidefinite matrices. *SIAM Journal on Optimization*, 5(1): 100–113, 1995. doi: 10.1137/0805005.

[11] A. Wächter and L. T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106(1): 25–57, 2006. doi: 10.1007/s10107-004-0559-y.

Thank you!



https://JuliaSmoothOptimizers.github.io