Practical Aspects of Primal-Dual Nonlinear Rescaling Method with Dynamic Scaling Parameter Update

Richard ANDRASIK

Department of Mathematical Analysis and Applications of Mathematics,
Faculty of Science, Palacky University Olomouc,
st. 17. listopadu 12, 771 46 Olomouc, Czech Republic

andrasik.richard@gmail.com

Abstract. *Primal-dual nonlinear rescaling method with dynamic scaling parameter update (PDNRD) is an optimization method from a class of nonlinear rescaling techniques. Previous work does not discuss practical aspects of PDNRD method such as the explanation and the setting of the parameters. To complete this framework, the parameters were described. Moreover, PDNRD method was applied on two quadratic programming problems with quadratic constraints and recommendations about the setting of the parameters were made.*

Keywords

Convex optimization, dynamic scaling parameter update, nonlinear rescaling method, parameters setting.

1. Introduction

The optimization theory is developing in parallel with the appearance of real-life problems and with the need to solve them. Thanks to that, some types of problems gained a special status among the others and many tools to solve these problems have been developed. An example is the linear programming (LP). In 1947, George Dantzig introduced the simplex method for solving LP [3]. However, the simplex method has a strong competitor – interior-point methods, especially primal-dual interior-point methods [15].

In the field of the nonlinear programming (NLP), interior-point methods have been also applied. However, the situation in NLP is more complicated in comparison with LP calculations and interior point methods are sometimes experiencing numerical difficulties.

This fact motivated Roman Polyak and Igor Griva to design an alternative method based on the nonlinear rescaling (NR) theory.

Nowadays, NR methods can solve large-scale NLP problems with thousands of variables and constraints. They were successfully used to the radiotherapy treatment planning and are applied at some hospitals in USA and Europe [1].

The basic idea of NR methods is a nonlinear transformation of constraint functions to improve the properties of Lagrangian. Originally, the modified barrier methods [8] were introduced along with few modified barrier functions. Afterwards, the log-sigmoid function was also considered usable for NR [9], [10]. Consequently, the pieces of knowledge were refined, put together and a generalization of these techniques led to the concept of NR methods and NR functions. Similar to progress with interior-point methods, the primal-dual nonlinear rescaling (PDNR) method was developed [11].

PDNR method is locally convergent with the Q-linear convergence rate. To improve these properties, PDNR method can be combined with another optimization method (e.g. the primal-dual path-following method) to obtain the global convergence [11]. Another way how to improve the convergence of PDNR method is a dynamic scaling parameter update [5] together with some globalization strategy (e.g. a step length computation). Recently, generalizations and other improvements [12], [13] were developed to improve the asymptotic convergence rate and to reduce the computational effort. The main purposes of this paper are to describe in detail parameters of PDNRD method and to give recommendations about their setting.

The paper is organized as follows. First, the convex optimization problem is stated and basic assumptions

are discussed. Then, NR functions are defined and the key idea of NR method is explained. Afterwards, basic primal-dual variant of NR method is presented. Next, PDNRD method is explained and its parameters are described. Finally, numerical experiments with different parameter settings were made and the results are presented in Section 7.

2.   Statement of the Problem

We consider the convex optimization problem

$$\begin{cases} \text{minimize} & f(x), \quad x \in \Re^n, \\ \text{subject to} & c_i(x) \geq 0, \quad i = 1, \dots, r. \end{cases} \quad (1)$$

Function f is convex and functions c_i are concave, $\forall i = 1, \dots, r$. Let $S \subseteq \Re^n$ be the admissible set of problem Eq. (1). For simplicity we define mapping $c : \Re^n \to \Re^r$ as

$$c(x) = (c_1(x), c_2(x), \dots, c_r(x))^{\mathrm{T}}, \quad \forall x \in \Re^n.$$

We suppose that:

- Functions f, c_i, $\forall i = 1, \dots, r$, are at least twice continuously differentiable on the set $\Re^n$.

- The optimal set $X^* = \operatorname{Argmin}\{f(x); x \in S\}$ is bounded and not empty.

- The Slater condition holds.

For problem Eq. (1) we define the Lagrangian

$$L(x; \lambda) = f(x) - \sum_{i=1}^{r} \lambda_i c_i(x). \quad (2)$$

Due to assumption (C), Karush-Kuhn-Tucker's (KKT) conditions can be used to test the optimality. If $\hat{x} \in X^*$ then there is a vector $\hat{\lambda} \in \Re^r$ such that

$$\begin{aligned} \nabla_x L(\hat{x}; \hat{\lambda}) &= 0, \\ \hat{\lambda} &\geq 0, \\ \hat{\lambda}_i c_i(\hat{x}) &= 0, \ \forall i = 1, \dots, r. \end{aligned} \quad (3)$$

Conversely, if a pair $(\hat{x}, \hat{\lambda}) \in S \times \Re^r$ satisfies Eq. (3) then $\hat{x} \in X^*$.

3.   Nonlinear Rescaling Functions

First, we define functions that will be used to transform constraints of problem Eq. (1).

Definition 1. Twice continuously differentiable function $\psi : (t_0; +\infty) \to \Re$, where $-\infty < t_0 < 0$, satisfying conditions

(i) $\quad \psi(0) = 0, \psi'(0) = 1,$
(ii) $\quad \psi'(t) > 0, \forall t \in (t_0; +\infty),$
(iii) $\quad \psi''(t) < 0, \forall t \in (t_0; +\infty),$
(iv) $\quad \exists a > 0 : \psi(t) \leq -at^2, \forall t \in (t_0; 0),$
(v) $\quad \exists b > 0 : \psi'(t) \leq bt^{-1}, \forall t > 0,$
(vi) $\quad \exists c > 0 : \psi''(t) \geq -ct^{-2}, \forall t > 0$

is called NR function.

Remark 1. It follows from (ii) and (iii) that NR function ψ is increasing and concave on the whole domain. Because of (ii) and (v) it is true that $\lim_{t \to +\infty} \psi'(t) = 0$. Similarly from (iii) and (vi) it holds that $\lim_{t \to +\infty} \psi''(t) = 0$.

For example, the exponential transformation, the modified logarithmic function and the hyperbolic barrier function defined by the following formulas

$$\begin{aligned} \psi_1(t) &= 1 - e^{-t}, \\ \psi_2(t) &= \ln(t + 1), \\ \psi_3(t) &= \frac{t}{t+1}. \end{aligned}$$
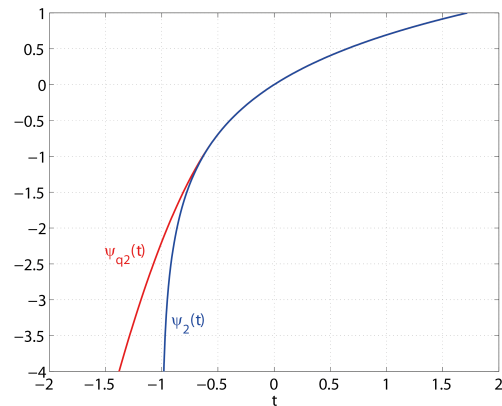
are nonlinear rescaling functions [11].

Fig. 1: Graphs of functions ψ_2 (blue) and ψ_{q_2} (red), $\tau = -\frac{1}{2}$.

Functions ψ_i, $i = 2, 3$, can be modified so that $\psi_i \in \mathcal{C}^2(\Re)$. The function ψ_1 is already twice continuously differentiable and the following modification is not necessary, of course it can be done. For a given parameter $\tau \in (-1; 0)$ the quadratic extrapolation is defined by the relations

$$\psi_{q_i}(t) = \begin{cases} \psi_i(t), & \text{for } t \geq \tau, \\ q_i(t) = A_i t^2 + B_i t + C_i, & \text{for } t < \tau, \end{cases}$$

whereas coefficients of the function q_i can be determined from the formulas

$$\psi_i(\tau) = q_i(\tau), \quad \psi_i'(\tau) = q_i'(\tau), \quad \psi_i''(\tau) = q_i''(\tau).$$

From this equations we obtain

$$
\begin{array}{rcl}
A_i & = & \frac{1}{2}\psi_i''(\tau), \\
B_i & = & \psi_i'(\tau) - \tau\psi_i''(\tau), \\
C_i & = & \psi_i(\tau) - \tau\psi_i'(\tau) + \frac{1}{2}\tau^2\psi_i''(\tau).
\end{array}
$$

4. Nonlinear Rescaling Approach

NR methods are based on the idea to transform problem Eq. (1) using a nonlinear rescaling function ψ defined on the whole real axis to the equivalent problem

$$
\begin{cases}
\text{minimize} & f(x), \quad x \in \Re^n, \\
\text{subject to} & k^{-1}\psi(kc_i(x)) \geq 0, \, i = 1, \ldots, r.
\end{cases}
\tag{4}
$$

From the definition of NR function, it is obvious that problems Eq. (1) and Eq. (4) have the same admissible sets, because

$$
\psi(kc_i(x)) \geq 0 \Leftrightarrow c_i(x) \geq 0, \; \forall i = 1, \ldots, r.
$$

Hence, the optimal sets are also the same. Positive real number k is the scaling parameter, which can be fixed or dynamically enlarged during the iteration process, see Algorithm 3.

The Lagrangian for equivalent problem Eq. (4) is given by the following formula:

$$
\mathcal{L}(x; \lambda, k) = f(x) - k^{-1}\sum_{i=1}^{r}\lambda_i\psi(kc_i(x)).
\tag{5}
$$

For any $r \in \mathbb{N}$ we denote

$$
\begin{array}{rcl}
\Re_+^r & = & \{v \in \Re^r; \; v_i \geq 0\}, \\
\Re_{++}^r & = & \{v \in \Re^r; \; v_i > 0\}.
\end{array}
$$

Suppose for a while that we know the solution of the dual problem $\lambda^* \in \Re_+^r$. Then it is sufficient to minimize the function $\mathcal{L}(x; \lambda^*, k)$ in the primal variable x. The constrained optimization problem would be transformed to an unconstrained optimization problem.

Since the Lagrange multipliers λ^* are not known, we estimate them and update them in every step of the method - just like the primal problem. In consequence, the constrained optimization problem is converted to a sequence of unconstrained optimization problems. Newton's method or its variant is applied in each step to minimize the Lagrangian $\mathcal{L}$ in the primal variable.

Algorithm 1. (The basic concept of NR methods)

Let $k > 0$ be a scaling parameter. Initial approximations $x^0 \in \Re^n$ and $\lambda^0 \in \Re_{++}^r$ are given. We suppose that an approximation $(x^s, \lambda^s) \in \Re^n \times \Re_{++}^r$, $s \in \mathbb{N}_0$,

is known already. We find the next primal-dual pair (x^{s+1}, λ^{s+1}) using the following formulas

$$
\begin{array}{rcl}
x^{s+1} & : & \nabla_x\mathcal{L}(x^{s+1}; \lambda^s, k) = 0, \\
\lambda_i^{s+1} & = & \psi'\left(kc_i(x^{s+1})\right)\lambda_i^s, \quad i = 1, \ldots, r.
\end{array}
\tag{6}
$$

The NR method converges for any fixed but large enough $k > 0$ under the standard second order optimality conditions [5].

Remark 2. If $\lambda^s \in \Re_{++}^r$, then also $\lambda^{s+1} \in \Re_{++}^r$. This property follows from condition (ii) of NR function ψ. In other words, NR methods are interior-point methods in the dual variable.

The Algorithm 1 is well defined due to the following theorem.

Theorem 1. *Suppose that X^* is bounded. For any given $(\lambda, k) \in \Re_{++}^r \times \Re_{++}$ there exists one and only one $\hat{x} \in \Re^n$ such that*

$$
\mathcal{L}(\hat{x}; \lambda, k) = \min_{x \in \Re^n} \mathcal{L}(x; \lambda, k).
$$

Proof: [10], page 206.

The main purpose of NR is to improve properties of the Lagrangian. The classical Lagrangian L – as a connection between the constrained and the unconstrained optimization – does not always work, because the existence of the unconstrained Lagrange minimizer is unknown in general. On the other hand, the unconstrained minimizer of the Lagrangian $\mathcal{L}$ always exists (according to Theorem 1). Moreover, NR dramatically sharpens the reaction of the Lagrangian to the constraint violation, which has an impact on the computations.

In comparison to interior-point methods, there is no "infinite wall". NR methods are exterior point methods. Also, the unbounded increase of the scaling parameter is not needed to guarantee the convergence of the method.

We illustrate the influence of NR on the Lagrangian on the following simple example.

Example 1. The Lagrangian of the following problem

$$
\begin{cases}
\text{minimize} & x, \quad x \in \Re^1, \\
\text{subject to} & x + 1 \geq 0,
\end{cases}
\tag{7}
$$
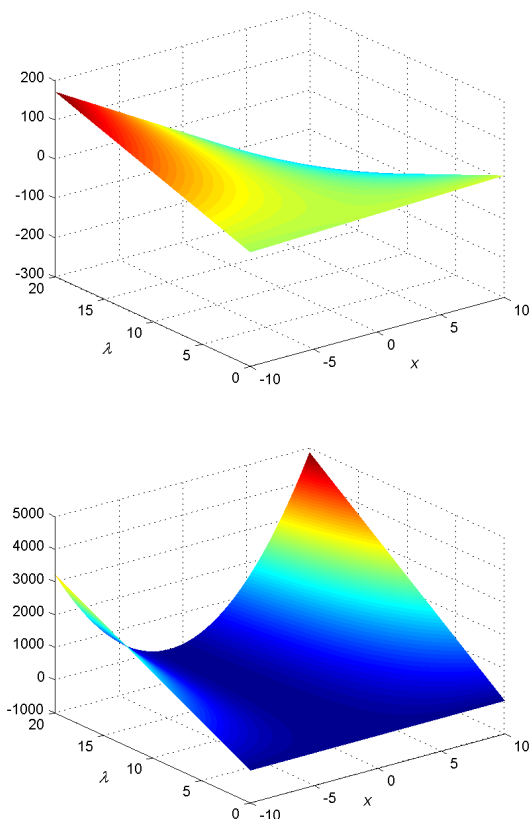
and the Lagrangian of the equivalent problem are considered.

$$
\begin{cases}
\text{minimize} & x, \quad x \in \Re^1, \\
\text{subject to} & k^{-1}\psi_{q_2}\left(k(x+1)\right) \geq 0,
\end{cases}
\tag{8}
$$

We take a look on the graphs of both Lagrangians, which are defined by the formulas

$$
\begin{aligned}
L(x;\lambda) &= x - \lambda(x+1), \\
\mathcal{L}(x;\lambda,k) &= x - k^{-1}\lambda\psi_{q_2}\left(k(x+1)\right).
\end{aligned}
\tag{9}
$$

NR fundamentally changes the shape of the Lagrangian. The sharp reaction of the Lagrangian $\mathcal{L}$ to the constraint violation is obvious in Fig. 2.

Fig. 2: The Lagrangian for the problem Eq. (7) - on the top - and for the equivalent problem Eq. (8) with $k = 1$ - on the bottom.

We project the graphs of the both Lagrangians to the phase plane in Fig. 3. Suppose that we want to find Lagrange minimizer for $\lambda = 2$. After substitution, formulas Eq. (9) have the form:

$$
\begin{aligned}
L(x;2) &= -x - 2, \\
\mathcal{L}(x;1,k) &= x - k^{-1}2\psi_{q_2}\left(k(x+1)\right).
\end{aligned}
$$

The function $L(x;2)$ is a decreasing linear function. We cannot find the Lagrange minimizer of it, because it is unbounded below.

If we rearrange the first equation in formulas Eq. (9) in the following way:

$$
L(x;\lambda) = (1-\lambda)x - \lambda,
$$

it is apparent that for any $\lambda \neq 1$ we cannot find the Lagrange minimizer. On the other hand, we know from Theorem 1 that the Lagrangian $\mathcal{L}$ has a minimizer for any $\lambda > 0$. This is the main contribution of NR.
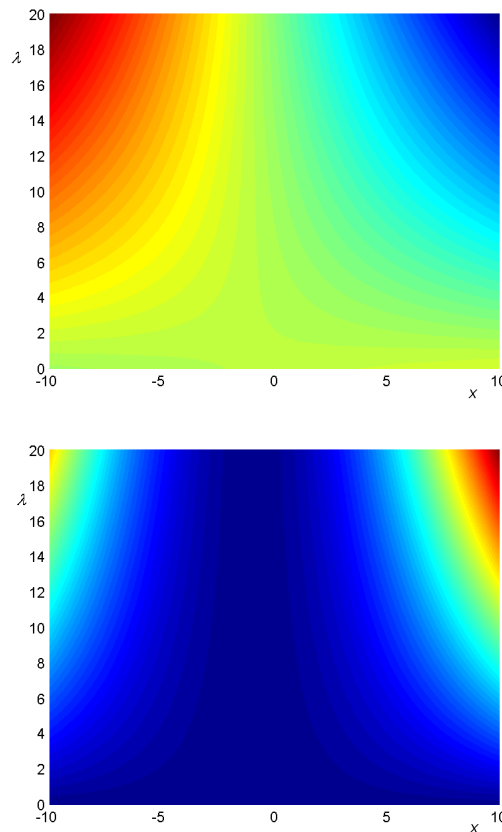
Fig. 3: The projection of the Lagrangian for the problem Eq. (7) - on the top - and for the equivalent problem Eq. (8) with $k = 1$ - on the bottom - to the phase plane.

5. Primal-dual Nonlinear Rescaling Method

The update of the Lagrange multipliers is clear from equations Eq. (6). When we want to calculate x^{s+1}, we must solve a nonlinear system of equations using Newton's method. Although this approach works well far from the solution, primal-dual variant of the NR method performs better in the neighborhood of the solution [5]. In Section 6, we will combine these two techniques to obtain a stable and globally convergent method.

Let approximations $x \in \Re^n$, $\lambda \in \Re^r_{++}$ be known. We suppose that

$$
\begin{aligned}
\hat{x} &= x + \Delta x, \\
\hat{\lambda} &= \bar{\lambda} + \Delta\lambda,
\end{aligned}
\tag{10}
$$

where $(\Delta x, \Delta \lambda)$ is a primal-dual Newton step and $\bar{\lambda}$ is a predictor of the Lagrange multipliers, which is given by the following formula

$$\bar{\lambda} = \Psi'(kc(x))\lambda,$$

where

$$\Psi'(kc(x)) = \operatorname{diag}(kc_i(x))_{i=1}^r.$$

The reason, why define $\bar{\lambda}$ in this way is hidden in the equality

$$\nabla_x \mathcal{L}(x; \lambda, k) = \nabla_x L(x; \bar{\lambda}).$$

One step of Newton's method consists of solving the system of $(n + r)$ linear equations

$$N(\cdot) \begin{bmatrix} \Delta x \\ \Delta \lambda \end{bmatrix} = \begin{bmatrix} -\nabla_x L(x; \bar{\lambda}) \\ 0 \end{bmatrix}, \qquad (11)$$

where a pair $(\Delta x, \Delta \lambda)$ is unknown and

$$N(\cdot) = \begin{bmatrix} \nabla_{xx}^2 L(x; \bar{\lambda}) & -\nabla c(x)^{\mathrm{T}} \\ -k\Psi''(kc(x))\Lambda\nabla c(x) & \mathbf{I}_r \end{bmatrix}.$$

Symbols $\Psi''(kc(x))$, Λ denotes diagonal matrices defined as

$$\begin{aligned} \Psi''(kc(x)) &= \operatorname{diag}(\psi''(kc_i(x)))_{i=1}^r, \\ \Lambda &= \operatorname{diag}(\lambda_i)_{i=1}^r. \end{aligned}$$

If $N(\cdot)$ is sparse, we can use numerical algebra techniques for sparse matrices [15] to solve the system Eq. (11).

In the opposite case, we express $\Delta \lambda$ from the second equation of the system Eq. (11) and we substitute it to the first equation. Now, we must solve only a system with n equations, instead of a system of $(n + r)$ equations. The system has the following form

$$M(\cdot)\Delta x = -\nabla_x L(x; \bar{\lambda}), \qquad (12)$$

where

$$M(\cdot) = \nabla_{xx}^2 L(x; \bar{\lambda}) - k\nabla c(x)^{\mathrm{T}}\Psi''(kc(x))\Lambda\nabla c(x).$$

Remark 3. It is obvious that $M(\cdot)$ is a symmetric matrix. It can be shown [11], page 120 that if the second order optimality conditions hold true at the point $(x^*; \lambda^*)$, then $M(\cdot)$ is a positive definite matrix for all $(x; \lambda)$ sufficiently close to the optimal primal-dual pair, when the scaling parameter is sufficiently large.

Algorithm 2. (One step of PDNR method)

Let $k \in \Re_{++}$ be fixed. The approximations $x \in \Re^n$, $\lambda \in \Re_{++}^r$ are given.

- We compute a dual predictor

$$\bar{\lambda} = \Psi'(kc(x))\lambda.$$

- We calculate the Newton step $(\Delta x, \Delta \lambda)$ from system Eq. (11), or

- we find out the primal Newton step Δx from formula Eq. (12) and then we compute the dual Newton step using the following relation

$$\Delta \lambda = k\Psi''(kc(x))\Lambda\nabla c(x)\Delta x.$$

- We calculate the new primal-dual approximation

$$x := x + \Delta x, \quad \lambda := \bar{\lambda} + \Delta \lambda.$$

The Algorithm 2 describes only one step of PDNR method. Although, the input vector of multipliers belongs to $\Re_{++}^r$, it is not guaranteed that the updated vector of multipliers has also all its components positive. In contrast, this statement is true for Algorithm 1, according to Remark 2. Therefore, PDNR method cannot stand alone in a general case and the improvement is needed (see Algorithm 3).

6. Dynamic Scaling Parameter Update

To obtain a higher convergence rate of the method, we dynamically change the scaling parameter. Moreover, we use Newton's method with a step length (e.g. the backtracking line search algorithm) to solve formulas Eq. (6).

We introduce a function which measures the distance between the approximation (x, λ) and the solution (x^*, λ^*).

Definition 2. The function $\nu : \Re^n \times \Re^r \to \Re_+$, defined as follows

$$\begin{aligned} \nu(x, \lambda) = \max \{ \; & \|\nabla_x L(x; \lambda)\|, \\ & -\min_{1 \le i \le r} c_i(x), \\ & -\min_{1 \le i \le r} \lambda_i, \\ & \textstyle\sum_{i=1}^r |\lambda_i c_i(x)| \; \}, \end{aligned} \qquad (13)$$

is called the merit function.

Formula (13) is motivated by KKT conditions. From the first order optimality conditions it follows that

$$\nu(\hat{x}, \hat{\lambda}) = 0 \Leftrightarrow (\hat{x}, \hat{\lambda}) \in X^*.$$

For known primal-dual pair (x, λ) we set the parameter k according to the relation

$$k = \nu(x, \lambda)^{-1/2}. \qquad (14)$$

It is obvious that if a primal-dual sequence $\{(x^s, \lambda^s)\}_0^{+\infty}$ tends to (x^*, λ^*), then $\nu(x^s, \lambda^s) \to 0^+$ and also $k_s \to +\infty$ when $s \to +\infty$.

To prevent from a singularity in system Eq. (11), we solve the following regularized system

$$N_k(\cdot) \begin{bmatrix} \Delta x \\ \Delta \lambda \end{bmatrix} = \begin{bmatrix} -\nabla_x L(x; \bar{\lambda}) \\ 0 \end{bmatrix}, \qquad (15)$$

where the primal-dual pair (x, λ) is an approximation of solution, k is the scaling parameter, $\bar{\lambda} = \Psi'(kc(x))\lambda$ is the dual predictor of the Lagrange multipliers and

$$N_k(\cdot) = \begin{bmatrix} \nabla_{xx}^2 L(x; \bar{\lambda}) + \frac{1}{k^2}\mathbf{I}_n & -\nabla c(x)^{\mathrm{T}} \\ -k\Psi''(kc(x))\Lambda\nabla c(x) & \mathbf{I}_r \end{bmatrix}.$$

It is important to remark that the regularization has no effect to the 1.5-superlinear convergence rate of PDNRD method [5].

In a comparison to interior-point methods, the inverse matrix to $N_k(\cdot)$ exists at the optimal point for every $k \in \Re_{++}$.

Assume that the approximation (x, λ) of the point (x^*, λ^*) is known. First, we use PDNRD method. If there is a superlinear decrease of the merit function, we have found the next approximation. In the opposite case, we use the primal Newton step Δx (calculated during the use of PDNRD method) to minimize the function $\mathcal{L}(x; \lambda, k)$, where λ and k are fixed. We apply the backtracking line search method to guarantee the global convergence in the minimization process. In this way, we obtained the globally convergent PDNRD method with the 1.5-superlinear convergence rate.

Algorithm 3. (The globally convergent PDNRD method)

An initial approximation $x^0 \in \Re^n$ is given. An accuracy parameter $\varepsilon > 0$ and an initial scaling parameter $k \in \Re_{++}$ are given. Parameters $q \in (0; 1)$, $\eta \in (0; 0.5)$, $\omega > 1$, $\sigma > 0$ and $\theta > 0$ are also given. Set $x := x^0$, $\lambda := (1, 1, \ldots, 1) \in \Re^r$, $\lambda_g := \lambda$ and $H := \nu(x, \lambda)$.

- If $H \leq \varepsilon$, then stop, output (x, λ).

- Find $\bar{\lambda}$ and $(\Delta x, \Delta \lambda)$ from primal-dual system Eq. (15) with known (x, λ) and set

$$\hat{x} := x + \Delta x, \ \hat{\lambda} := \bar{\lambda} + \Delta\lambda, \ \hat{H} := \nu(\hat{x}, \hat{\lambda}).$$

- If $\hat{H} \leq \min\left\{H^{3/2-\theta}, 1-\theta\right\}$, then set

$$x := \hat{x}, \ \lambda := \hat{\lambda}, \ H := \hat{H}, \ k := \max\left\{\frac{1}{\sqrt{H}}, k\right\}$$

and go to step 1.

- Find $\alpha \in (0; 1\rangle$ so that it holds

$$\mathcal{L}(x + \alpha\Delta x; \lambda_g, k) - \mathcal{L}(x; \lambda_g, k)$$
$$\leq \eta\alpha\Delta x^{\mathrm{T}}\nabla_x\mathcal{L}(x; \lambda_g, k),$$

using the backtracking line search algorithm.

- Set

$$x := x + \alpha\Delta x, \ \hat{\lambda} := \Psi'(kc(x))\lambda_g.$$

- If

$$\|\nabla_x\mathcal{L}(x; \lambda_g, k)\| \leq \frac{\sigma}{k}\|\hat{\lambda} - \lambda_g\|,$$

then go to step 8.

- Find $(\Delta x, \Delta \lambda)$ from primal-dual system Eq. (15) with known (x, λ_g) and go to step 4.

- If $\nu(x, \hat{\lambda}) \leq qH$, then set

$$\lambda := \hat{\lambda}, \ \lambda_g := \hat{\lambda}, \ H := \nu(x, \hat{\lambda}), \ k := \max\left\{\frac{1}{\sqrt{H}}, k\right\}$$

and go to step 1.

- Set $k := \omega k$ and go to step 7.

7. Numerical Experiments

From the finite element approximation of contact problems of the linear elasticity with the friction in three space dimensions arise a minimization problem

$$\begin{cases} \text{minimize} & \frac{1}{2}x^{\mathrm{T}}Ax - x^{\mathrm{T}}b, \quad x \in \Re^n \\ \text{subject to} & g_i^2 - x_{i+m}^2 - x_{i+2m}^2 \geq 0, \ i \in \mathcal{I}, \\ & x_i - l_i \geq 0, \ i \in \mathcal{I}, \end{cases} \quad (16)$$

where $n = 3m$ is the number of variables, $A \in \Re^{n \times n}$ is a symmetric and positive definite matrix, $b \in \Re^n$, $g \in \Re_+^m$, $l \in \Re^m$ and $\mathcal{I} = \{1, 2, \ldots, m\}$. This is a convex programming problem so we can use NR approach to solve it. We use the function ψ_{q_2} with $\tau = \frac{1}{2}$ to rescale the conditions and subsequently we obtain the equivalent problem. The reason for this choice is that for a sufficiently large class of functions $c(x)$ is the function $-\ln c(x)$ self-concordant. If we apply Newton's method to a self-concordant function, we can say something more about the convergence of Newton's method [2]).

PDNRD method was tested on two model problems - the chord problem, which in contrary to Eq. (16) contains also unconstrained variables, and the steel brick problem, whose finite element approximation leads exactly to Eq. (16). All computations were performed in MATLAB on the PC Intel Core i7 (2.4 GHz) with 8 GB RAM. In tables below we reported the number of iterations (it), the number of solutions of the primal-dual system (n_S) and the solution time in seconds $(time)$. If some data is missing, it means that the solution time was too long in the comparison to the other cases in the table.

7.1. Chord Problem

We consider a problem

$$\min_{u \in \mathcal{K}} \mathcal{J}(u), \qquad (17)$$

where

$$\mathcal{J}(u) = \frac{1}{2} \int_0^1 \|u'(t)\|^2 \mathrm{d}t - \int_0^1 u(t)^{\mathrm{T}} f(t) \mathrm{d}t,$$

$$\mathcal{K} = \{ \quad u \in \left(H_0^1(0;1) \right)^2 : \ u_2(t) \geq 0, \forall t \in (0;0.5),$$
$$\|u(t)\| \leq 1.4, \forall t \in (0.5;1) \quad \},$$

$$f(t) = \left(36\pi^2 \sin 6\pi t, -4\pi^2 \sin 2\pi t \right)^{\mathrm{T}}.$$
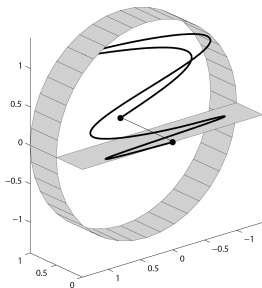

Fig. 4: The chord deformation.

Minimization problem Eq. (17) describes a loaded chord fixed at the endpoints. The chord is partially above a plain and partially inside a cylindrical tube Fig. 4. The function $u(t)$ is the chord deflection. The chord problem was presented as a model problem in [6].

The objective function $\frac{1}{2} x^{\mathrm{T}} \mathrm{A} x - x^{\mathrm{T}} b$, $x \in \Re^n$ matches the convex quadratic functional $\mathcal{J}(u)$, linear constraints in problem Eq. (16) matches the constraint $u_2(t) \geq 0$, $\forall t \in (0;0.5)$ from the definition of the set $\mathcal{K}$ and quadratic constraints matches $\|u(t)\| \leq 1,4$, $\forall t \in (0.5;1)$.

First, we decided if the primal-dual system has a sparse matrix. We say that the matrix is sparse if it has at most 10 % nonzero elements. In Fig. 5, the structure of the matrix $N_k(\cdot) \in \Re^{192 \times 192}$ for $n = 128$ is depicted. This matrix has only 636 nonzero elements (1.7 %). Hence, $N_k(\cdot)$ is the sparse matrix. In the same way we could argue for other choices of n.

We solved the chord problem for different settings of parameters of PDNRD method. The main result from Tab. 1 is a non-increasing number of iterations and also a non-increasing number of solutions of the primal-dual system while increasing the number of variables. The best choice of the scaling parameter was $k_{init} = 10$ in
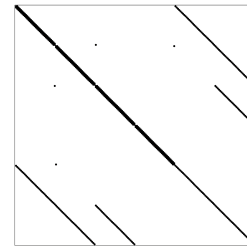

Fig. 5: The structure of the matrix $N_k(\cdot) \in \Re^{192 \times 192}$ (the chord problem). Black colour indicates 636 nonzero elements.

Tab. 1: The chord problem. PDNRD method with parameters $\omega = 10$, $\sigma = \frac{1}{2} k_{init}$, $\theta = 0.4$, $q = 0.5$, $\eta = 0.01$, $\varepsilon = 10^{-6}$.

		k_{init}		
n	r	10	10^2	10^3
64	32	10/31/0.173	4/23/0.118	7/7/0.021
128	64	7/50/0.238	6/33/0.126	3/79/0.351
256	128	8/60/0.515	8/37/0.454	3/143/1.099
512	256	11/65/1.989	5/49/2.037	7/150/5.036
1024	512	12/70/10.330	6/72/13.088	8/100/22.229
2048	1024	12/58/37.640	6/70/53.256	7/66/50.650
4096	2048	13/74/221.834	—	—
		$it/n_S/time$	$it/n_S/time$	$it/n_S/time$

Tab. 2: The chord problem ($n = 1024$). PDNRD with parameters $\omega = 10$, $\sigma = \frac{1}{2} k_{init}$, $k_{init} = 10$, $\varepsilon = 10^{-6}$.

	$\eta = 0.01$	
q	$\theta = 0.1$	$\theta = 0.4$
0.1	6/49/8.015	6/49/8.054
0.5	10/65/11.554	12/70/10.518
0.9	30/99/16.347	31/77/13.548
	$\eta = 0.3$	
q	$\theta = 0.1$	$\theta = 0.4$
0.1	6/49/8.133	6/49/8.036
0.5	10/68/11.503	12/71/10.671
0.9	30/92/15.094	31/78/12.885
	$it/n_S/time$	$it/n_S/time$

this case, because of the best performance for larger problems.

From Tab. 2 it is obvious that the choice $q = 0.1$ was the best one. Parameters η and θ had no significant impact on the computation in this case. ∎

7.2. Steel Brick Problem

Let us consider a steel brick lying on the rigid obstacle. The brick occupies the domain $\mathcal{S} = (0;3) \times (0;1) \times$

$(0; 1)$. The boundary ∂S is divided into three parts

$$\begin{aligned}
\Gamma_u &= \{0\} \times (0; 1) \times (0; 1), \\
\Gamma_c &= (0; 3) \times (0; 1) \times \{0\}, \\
\Gamma_p &= \partial S \backslash (\bar{\Gamma}_u \cup \bar{\Gamma}_c),
\end{aligned}$$

on which different boundary conditions are given (see Fig. 6). The problem is described in the detail in [6].
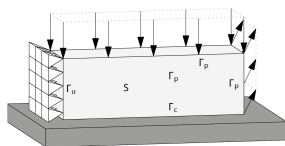
Fig. 6: The steel brick.

The progress was similar to the previous one considering the chord problem. First, the matrix $N_k(\cdot)$ was tested whether it is sparse or not. In Fig. 7, the structure of the matrix $N_k(\cdot) \in \Re^{150 \times 150}$ for $n = 90$ is depicted. This matrix contains 8340 nonzero elements (37 %). Hence, the matrix $N_k(\cdot)$ is not sparse and we solved the primal-dual system in a different way (see Section 5).
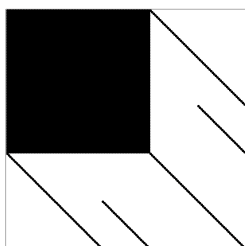
Fig. 7: The structure of matrix $N_k(\cdot) \in \Re^{150 \times 150}$ (steel brick problem). Black colour indicates 8340 nonzero elements.

The number of variables was denoted as $n = 3\ m$ and the number of conditions as $r = 2\ m$. The steel brick problem was solved for different choices of n. According to the data in Tab. 3, the best setting for the initial values of the scaling parameter was $k_{init} = 10$ in this problem. Number n_S rose with increasing k_{init}.

The results from Tab. 4 supports that the choice $\eta = 0.01$ is better than $\eta = 0.3$. The parameter θ has no significant impact on the computation.

Tab. 3: The steel brick problem. PDNRD method with parameters $\omega = 10$, $\sigma = \frac{1}{2}k_{init}$, $\theta = 0.4$, $q = 0.5$, $\eta = 0.01$, $\varepsilon = 10^{-6}$.

		k_{init}		
n	r	10	10^2	10^3
54	36	6/39/0.153	5/54/0.268	4/119/0.570
90	60	6/38/0.218	6/45/0.239	4/102/0.743
180	120	8/40/0.257	4/90/0.700	4/136/1.653
324	216	7/37/0.663	10/35/0.619	4/130/3.146
648	432	12/67/7.258	10/77/10.003	6/123/15.696
		$it/n_S/time$	$it/n_S/time$	$it/n_S/time$

Tab. 4: The steel brick problem ($n = 648$). PDNRD method with parameters $\omega = 10$, $\sigma = \frac{1}{2}k_{init}$, $k_{init} = 10$, $\varepsilon = 10^{-6}$.

	$\eta = 0.01$	
q	$\theta = 0.1$	$\theta = 0.4$
0.1	5/46/5.562	5/46/5.767
0.5	12/45/4.460	12/67/7.135
0.9	12/67/7.545	12/67/7.105
	$\eta = 0.3$	
q	$\theta = 0.1$	$\theta = 0.4$
0.1	5/44/5.593	5/44/5.482
0.5	11/41/4.398	12/62/8.820
0.9	12/62/8.761	12/62/8.718
	$it/n_S/time$	$it/n_S/time$

The characteristic property of PDNRD method is the "hot" start of the method (as it was observed and proved in [11]). "Hot" start means that from some point only one Newton step is needed to sufficiently shrink the distance between a current approximation and the solution. We can expect that a computational effort to solve the problem with the accuracy 10^{-12} will not be dramatically higher than with an accuracy 10^{-6}. Our results support this statement (compare Tab. 3 and Tab. 5).

Tab. 5: The steel brick problem. PDNRD method with parameters $\omega = 10$, $\sigma = \frac{1}{2}k_{init}$, $k_{init} = 10$, $\theta = 0.4$, $q = 0.5$, $\eta = 0.01$, $\varepsilon = 10^{-12}$.

n	r	$k_{init} = 10$
54	36	8/45/0.165
90	60	10/51/0.252
180	120	12/44/0.269
324	216	10/47/0.806
648	432	14/71/7.371
		$it/n_S/time$

7.3. Parameters

When using a numerical method the setting of the parameters is very important. Experience with numerical experiments helps us to find the optimal setting. Based

on the tests of PDNRD method on examples from Sections 7.1 and 7.2, we made considerations about a suitable setting of the parameters.

Factor ω

The factor ω affects the rate of the increase of the scaling parameter. For $\omega \in \langle 5; 20 \rangle$ we obtain almost the same results. So we set $\omega = 10$. However, even for the other choices $\omega > 1$ there are not any significant changes. At most, it may happen that it takes a few extra steps of the method.

Parameter σ

The choice of the parameter σ is less important than the choice of the ratio between σ and the initial choice of the scaling parameter k_{init}. It is the fraction $\frac{\sigma}{k}$ which decides about the number of inner iterations (and thus about the number of Newton steps in the damped phase of Newton's method). It is clear that for $\sigma >> k$ there are too little inner steps. On the other hand, for $\sigma << k$ there is too many of them. According to the results of the numerical experiments, the choice $\sigma = \frac{1}{2} k_{init}$ is suitable.

Parameter θ

The parameter θ affects whether the inner solver runs or not. If we set $\theta = 0$, the inner solver runs whenever the 1.5-superlinear decrease of the merit function was not achieved. Choosing $\theta = 0.5$, we are saying that we are satisfied with only linear decrease of the merit function. Therefore, it is wise to set $\theta \in \langle 0; 0.5 \rangle$. It appears (see Tab. 2 and Tab. 4) that the method does not depend on this parameter.

In the condition, which is related to the parameter θ, the term $\min \left\{ H^{3/2 - \theta}, 1 - \theta \right\}$ is calculated. We can ask why do not simply use the term $H^{3/2 - \theta}$ instead of the previous one. As we know, the classical Newton method is effective only in the neighborhood of the solution, so if we are "far" from the solution the damped Newton method is better suited to use. The expression "far" means that the merit function is greater than $1 - \theta$.

Factor q

The number of inner steps is influenced by the fraction $\frac{\sigma}{k}$ together with the factor q. This factor also affects how often is the scaling parameter increased. Due to the way in which is this parameter used in PDNRD method, it is needed to set $q \in (0; 1)$. Moreover, we must choose the factor q so that the method does not use too many inner steps. Based on the data in Tab. 2 and Tab. 4, any factor $q \in (0; 1)$ was suitable. However, the best results were obtained by setting $q = 0.1$.

Parameter η

The backtracking line search parameter η is usually chosen in the range from 0.01 to 0.3 (see [2] page 466). According to data in Tab. 2 and Tab. 4 the choice of the parameter $\eta \in \langle 0.01; 0.3 \rangle$ is arbitrary.

Initial value of the scaling parameter

Based on the results shown in Tab. 1 and Tab. 3, increasing the initial value of the scaling led to instabilities in computations. The best results in the both model examples were obtained with $k_{init} = 10$. Therefore, we recommend setting k_{init} around ten or less. Further reason to set k_{init} rather lower is that the scaling parameter can be increased during the computation, but it cannot be decreased.

Initial approximation

The computation also depends on the initial approximation. The number of steps can be decreased, if "lucky" initial approximation is chosen. In problems like a beam deflection or contact problems we choose an initial state of the system as the initial approximation, because the shape changes of a body are usually very small. Thus we set $x^0 = (0, 0, \ldots, 0)$.

8.   Conclusion

The practical aspects of PDNRD method were described. Especially, the meaning of the parameters of PDNRD method were explained in detail. PDNRD method was tested on two quadratic programming problems with quadratic constraints (the chord problem and the steel brick problem). Based on the tests, the recommendations about setting the parameters were made.

The solution of the primal-dual system is the most expensive operation, thus the number of solutions of the primal-dual system determines the total complexity of computations. It was found out that the increasing number of variables in both presented problems has not a consequence in the increasing number of solutions of the primal-dual system (see Tab. 1 and Tab. 3). This fact supports the applicability of PDNRD method on problems of an arbitrary size.

Acknowledgment

References

[1] ALBER, M. and R. REEMTSEN. Intensity modulated radiotherapy treatment planning by use of a barrier-penalty multiplier method. *Optimization Methods and Software*. 2007, vol. 22, iss. 3, pp. 391–411. ISSN 1055-6788. DOI: 10.1080/10556780600604940.

[2] BOYD, S. P. and L. VANDENBERGHE. *Convex optimization*. Cambridge: Cambridge University Press, 2004. ISBN 978-0-521-83378-3.

[3] DANTZIG, G. B. and L. VANDENBERGHE. *Linear programming and extensions*. Princeton, N.J.: Princeton University Press, 1998. ISBN 978-069-1059-136.

[4] GRIVA, I., S. G. NASH and A. SOFER. *Linear and nonlinear optimization*. 2nd ed. Philadephia: Society for Industrial and Applied Mathematics, 2009. ISBN 978-0-89871-661-0.

[5] GRIVA, I. and R. A. POLYAK. Primal-dual nonlinear rescaling method with dynamic scaling parameter update. *Mathematical Programming*. 2006, vol. 106, iss 2, pp. 237–259. ISSN 1436-4646. DOI: 10.1007/s10107-005-0603-6.

[6] KUCERA, R., J. MACHALOVA, H. NETUKA and P. ZENCAK. An interior-point algorithm for the minimization arising from 3D contact problems with friction. *Optimization Methods and Software*. 2013, vol. 28, iss. 6, pp. 1195–1217. ISSN 1029-4937. DOI: 10.1080/10556788.2012.684352.

[7] NOCEDAL, J. and S. J. WRIGHT. *Numerical optimization*. 2nd ed. New York: Springer, 2006, Springer series in operations research and financial engineering. ISBN 03-873-0303-0.

[8] POLYAK, R. A. Modified barrier functions (theory and methods). *Mathematical Programming*. 1992, vol. 54, iss. 1–3, pp. 177–222. ISSN 1436-4646. DOI: 10.1007/BF01586050.

[9] POLYAK R. A. Log-Sigmoid Multipliers Method in Constrained Optimization. *Annals of Opereration Research*. 2001, vol. 101, iss. 1–4, pp. 427–460. ISSN 1572-9338. DOI: 0.1023/A:1010938423538.

[10] POLYAK, R. A. Nonlinear rescaling vs. smoothing technique in convex optimization. *Mathematical Programming*. 2002, vol. 92, iss. 2, pp. 197–235. ISSN 1436-4646. DOI: 10.1007/s101070100293.

[11] POLYAK, R. A. Primal–dual exterior point method for convex optimization. *Optimization Methods and Software*. 2008, vol. 23, iss. 1, pp. 141–160. ISSN 1573-2878. DOI: 10.1080/10556780701363065.

[12] GRIVA, I. and R. A. POLYAK. 1.5-Q-superlinear convergence of an exterior-point method for constrained optimization. *Journal of Global Optimization*. 2008, vol. 40, iss. 4, pp. 679–695. ISSN 1573-2916. DOI: 10.1007/s10898-006-9117-x.

[13] POLYAK, R. A. and I. GRIVA. Proximal point nonlinear rescaling method for convex optimization. *Numerical Algebra, Control and Optimization*. 2011, vol. 1, iss. 2, pp. 283–299. ISSN 2155-3297. DOI: 10.3934/naco.2011.1.283.

[14] POLYAK, R. A. and M. TEBOULLE. Nonlinear rescaling and proximal-like methods in convex optimization. *Mathematical Programming*. 1997, vol. 76, iss. 2, pp. 265–284. ISSN 1436-4646. DOI: 10.1007/BF02614440.

[15] WRIGHT, S. J. *Primal-dual interior-point methods*. Philadelphia: Society for Industrial and Applied Mathematics, 1997. ISBN 08-987-1382-X.

About Authors

Richard ANDRASIK was born in Karvina, Czech republic, 1988. He received his M.Sc. from Palacky University Olomouc in 2012. Currently he is a Ph.D. student. His research interests include mathematical and physical models of fluid dynamics, optimization theory and numerical methods for solving partial differential equations.