PRIMAL–DUAL METHODS FOR NONLINEAR CONSTRAINED OPTIMIZATION

IGOR GRIVA
Department of Mathematical Sciences, George Mason University, Fairfax, Virginia

ROMAN A. POLYAK
Department of SEOR and Mathematical Sciences, George Mason University, Fairfax, Virginia

HISTORICAL NOTE

In 1797 in his book "Theorie des functions analytiques" Joseph Luis Lagrange (1736–1813) introduced the Lagrangian and Lagrange multipliers rule for solving optimization problems with equality constraints.

> ... If a function of several variables should be a maximum or minimum and there are between these variables one or several equations, then it will be suffice to add to the proposed function the functions that should be zero, each multiplied by an undetermined quantity, and then to look for the maximum and the minimum as if the variables were independent; the equation that one will find combined with the given equations, will serve to determine all the unknowns.
>
> J.-L. Lagrange

The Lagrange multipliers rule is only a necessary but not sufficient condition for an equality constrained optimum. The primal–dual (PD) vector is neither a maximum nor a minimum of the Lagrangian: it is a saddle point. Nevertheless, for more than 200 years the Lagrangian has remained an invaluable tool for optimization. Moreover, with time, the great value of the Lagrangian has become increasingly evident. The Lagrangian is the main instrument for establishing optimality conditions, the basic ingredient in the Lagrange duality, and one of the most important tools in numerical constrained optimization. Any PD method for solving constrained optimization problems in one way or another uses the Lagrangian and the Lagrange multipliers.

OPTIMIZATION WITH EQUALITY CONSTRAINTS: LAGRANGE SYSTEM OF EQUATIONS

Consider $q+1$ smooth enough functions $f,\ g_j : \mathbb{R}^n \to \mathbb{R},\ j = 1, \ldots, q$ and the feasible set

$$\Omega = \left\{x : g_j(x) = 0, j = 1, \ldots, q\right\}.$$

The equality constrained optimization (ECO) problem consists of finding

$$(\mathcal{ECO}) \qquad f(x^*) = \min\{f(x)|x \in \Omega\}.$$

The Lagrangian $L : \mathbb{R}^n \times \mathbb{R}^q \to \mathbb{R}^1$ for $(\mathcal{ECO})$ is given by the formula

$$L(x, v) = f(x) - \sum_{j=1}^{q} v_j g_j(x).$$

Let us consider the vector function $g^T(x) = \left(g_1(x), \ldots, g_q(x)\right)$, the Jacobian $J(g(x)) = \nabla g(x) = \left(\nabla g_1(x), \ldots, \nabla g_q(x)\right)^T$ and assume

$$\text{rank}\, \nabla g(x^*) = q < n. \qquad (1)$$

Then for x^* to be a $(\mathcal{ECO})$ solution, it is necessary the existence of $v^* \in \mathbb{R}^q$ that the pair (x^*, v^*) is a solution to the following Lagrange system of equations:

$$\nabla_x L(x, v) = \nabla f(x) - \nabla g^T(x) v = 0, \qquad (2)$$

$$g_i(x) = 0, \quad i = 1, \ldots, q. \qquad (3)$$

We consider the Hessian

$$\nabla^2_{xx} L(x, v) = \nabla^2 f(x) - \sum_{i=1}^{q} v_i \nabla^2 g_i(x)$$

Wiley Encyclopedia of Operations Research and Management Science, edited by James J. Cochran

of the Lagrangian $L(x, v)$. The regularity condition (1), together with sufficient condition for the minimizer x^* to be isolated that is

$$\langle \nabla^2_{xx} L(x^*, v^*)\xi, \xi\rangle > \geq m\langle \xi, \xi\rangle, \quad \forall \xi : \nabla g(x^*) \quad \xi = 0,\ m > 0, \tag{4}$$

comprise the standard second-order optimality conditions for the $(\mathcal{ECO})$ problem.

Application of Newton's method to the nonlinear PD system of Equations (2), (3) leads to one of the first PD methods for constrained optimization [1].

By linearizing the system (2), (3) and ignoring terms of the second and higher order, one obtains the following linear PD system for finding the Newton direction $(\Delta x, \Delta v)$:

$$\begin{bmatrix} \nabla^2_{xx} L(\cdot) & -\nabla g^T(\cdot) \\ \nabla g(\cdot) & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta v \end{bmatrix} = \begin{bmatrix} -\nabla_x L(\cdot) \\ -g(\cdot) \end{bmatrix}. \tag{5}$$

Let $S(y^*, \delta) = \{y = (x, v) : \|y - y^*\| \leq \delta\}$, where $\|\|$ is the Eucledian norm, δ small enough, and $0 < \epsilon \ll \delta$ is the desired accuracy. The following merit function

$$\nu(y) = \nu(x, v) = \max\left\{ \|\nabla_x L(x, v)\|,\ \max_{1\leq i\leq q} |g_i(x)| \right\}, \tag{6}$$

measures the distance between the PD approximation (x, v) and the solution (x^*, v^*). If f and g_i are smooth enough and the second-order optimality conditions for $(\mathcal{ECO})$ are satisfied, then for $y \in S(y^*, \delta)$ we have

$$\nu(y) = 0 \Leftrightarrow y = y^*, \tag{7}$$

and there are $0 < m_0 < M_0$ that

$$m_0\|y - y^*\| \leq \nu(y) \leq M_0\|y - y^*\|. \tag{8}$$

The PD method for $(\mathcal{ECO})$ consists of the following operations:

PDECOM-Primal–Dual Equality Constrained Optimization Method

Step 0. Let $y \in S(y^*, \delta)$.
Step 1. If $\nu(y) \leq \epsilon$, Output y as the solution.
Step 2. Find Δx and Δv from (5).
Step 3. Update the primal–dual pair by the following formulas

$$x := x + \Delta x, \quad v := v + \Delta v.$$

Step 4. Goto Step 1.

If the standard second-order optimality for conditions for $(\mathcal{ECO})$ are satisfied, the Lipschitz conditions for the Hessians $\nabla^2 f(x)$, $\nabla^2 g_i, i = 1, \ldots, q$ hold, $y = (x, v) \in S(y^*, \delta)$ and $\delta > 0$ is small enough, then the PDECOM generates the PD sequence that converges to the PD solution $y^* = (x^*, v^*)$ with quadratic rate; that is, the following bound holds:

$$\|\hat{y} - y^*\| \leq C\|y - y^*\|^2, \tag{9}$$

where $C > 0$ is independent from $y \in S(y^*, \delta)$ and depends only on the problem data [1].

AUGMENTED LAGRANGIAN AND PRIMAL–DUAL AUGMENTED LAGRANGIAN METHOD

For a given $k > 0$, the augmented Lagrangian (AL) $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^q \times \mathbb{R}_+ \to \mathbb{R}^1$ is defined by the following formula [2,3]:

$$\mathcal{L}(x, v, k) = f(x) - \sum_{j=1}^{q} v_j g_j(x) + \frac{k}{2}\sum_{j=1}^{q} g_j^2(x). \tag{10}$$

The AL method alternates the unconstrained minimization of the AL $\mathcal{L}(x, v, k)$ in the primal space with a Lagrange multipliers update.

Let $\delta > 0$ be sufficiently small, $0 < \epsilon \ll \delta$ be the desired accuracy, $k > 0$ be sufficiently large, and $y = (x, v) \in S(y^*, \delta)$. One step of the classical AL method consists of finding the primal minimizer

$$\hat{x} = \arg\min_{x \in \mathbb{R}^n} \mathcal{L}(x, v, k), \tag{11}$$

followed by the Lagrange multipliers update.

In other words, for a given scaling parameter $k > 0$ and a starting point $y = (x, v) \in S(y^*, \delta)$, one step of the AL method is equivalent to solving the following nonlinear PD system for $\hat{x}$ and $\hat{v}$:

$$\nabla_x L(\hat{x}, \hat{v}) = \nabla f(\hat{x}) - \sum_{j=1}^{q} \hat{v}_j \nabla g_j(\hat{x}) = 0, \tag{12}$$

$$\hat{v} - v + kg(\hat{x}) = 0. \tag{13}$$

Application of Newton's method for solving the PD system (12)–(13) for $(\hat{x}, \hat{v})$ together with a proper update of the penalty parameter $k > 0$ leads to the primal–dual augmented Lagrangian (PDAL) method for solving $(\mathcal{ECO})$ problems [4].

By linearizing the system (12)–(13) and ignoring terms of the second and higher order, we obtain the following linear PD system for finding the Newton direction $(\Delta x, \Delta v)$:

$$\begin{bmatrix} \nabla^2_{xx} L(\cdot) & -\nabla g^T(\cdot) \\ \nabla g(\cdot) & \frac{1}{k} I_q \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta v \end{bmatrix} = \begin{bmatrix} -\nabla_x L(\cdot) \\ -g(\cdot) \end{bmatrix}, \tag{14}$$

where I_q is the identity matrices in $\mathbb{R}^q$. Note that if $k \to \infty$, then the system (14) gets close to the system (5). Therefore, by changing $k > 0$ properly it is possible to achieve a quadratic rate of convergence for the PDAL method.

PDALM - Primal–Dual Augmented Lagrangian Method

Step 0. Let $y \in S(y^*, \delta)$ and $k > 0$ is large enough.
Step 1. If $\nu(y) \leq \epsilon$, Output y as a solution.
Step 2. Set $k = \nu(x, v)^{-1}$.
Step 3. Find Δx and Δv from (14).
Step 4. Update the primal–dual pair by the formulas

$$x := x + \Delta x, \quad v := v + \Delta v.$$

Step 5. Goto Step 1.

If the standard second-order optimality conditions for $(\mathcal{ECO})$ are satisfied, the Lipschitz conditions for the Hessians $\nabla^2 f(x)$, $\nabla^2 g_i$, $i = 1, \ldots, q$ hold, $y = (x, v) \in S(y^*, \delta)$, and $\delta > 0$ is sufficiently small, then the PDALM generates a PD sequence that converges to the PD solution $y^* = (x^*, v^*)$ with a quadratic rate, that is Equation (9) holds [4].

OPTIMIZATION WITH INEQUALITY CONSTRAINTS

Let $f : \mathbb{R}^n \to \mathbb{R}^1$ be convex and all $c_i : \mathbb{R}^n \to \mathbb{R}^1$, $i = 1, \ldots, p$ be concave and smooth functions. Consider a convex set $\Omega = \{x \in \mathbb{R}^n : c_i(x) \geq 0, i = 1, \ldots, p\}$ and the following convex inequality constrained optimization (ICO) problem:

$$(\mathcal{ICO}) \qquad f(x^*) = \min\{f(x) | x \in \Omega\}.$$

Let us assume that

(a) The primal optimal set X^* is not empty and bounded.
(b) The Slater's condition holds; that is, there exists $\hat{x} \in \mathbb{R}^n : c_i(\hat{x}) > 0$, $i = 1, \ldots, p$.

Due to the assumption **B**, the Karush–Kuhn–Tucker's (KKT's) conditions hold true; that is, there exists a vector $u^* = (u_1^*, \ldots, u_p^*) \in \mathbb{R}_+^p$ such that

$$\nabla_x L(x^*, u^*) = \nabla f(x^*) - \sum_{i=1}^{p} u_i^* \nabla c_i(x^*) = 0, \quad (15)$$

$$c_i(x^*) \geq 0, \quad u_i^* \geq 0, \quad i = 1, \ldots, p, \quad (16)$$

and the complementary slackness conditions

$$u_i^* c_i(x^*) = 0, \, i = 1, \ldots, p \quad (17)$$

hold true.

Let $I = \{i : c_i(x^*) = 0\} = \{1, \ldots, r\}$ be the set of indices of the active at x^* constraints, $c_{(r)}^T(x) = \left(c_1(x), \ldots, c_r(x)\right)$ be the vector function of the active at x^* constraints and $\nabla c_{(r)}(x) = J\left(c_{(r)}(x)\right)$ be the corresponding Jacobian. The standard second-order optimality conditions for $(\mathcal{ICO})$ consists of existence $u^* \in \mathbb{R}_+^p$ and $m > 0$ such that

$$\operatorname{rank}\left(\nabla c_{(r)}(x^*)\right) = r, \, u_i^* > 0, i = 1, \ldots, r \quad (18)$$

and

$$\langle \nabla_{xx}^2 L(x^*, u^*)\xi, \xi \rangle \geq m\langle \xi, \xi \rangle, \, \forall \xi : \nabla c_{(r)}(x^*)\xi = 0. \quad (19)$$

The dual to $(\mathcal{ICO})$ problem consists of finding

$$(\mathcal{D}) \qquad d(u^*) = \max\left\{d(u) | u \in \mathbb{R}_+^p, \right\},$$

where $d(u) = \inf_{x \in \mathbb{R}^n} L(x, u)$ is the dual function.

For convex $(\mathcal{ICO})$ that satisfies the Slater condition we have

$$f(x^*) = d(u^*).$$

The PD methods generate PD sequences $\{x^s, u^s\}_{s=1}^{\infty}$ that

$$f(x^*) = \lim_{s \to \infty} f(x^s) = \lim_{s \to \infty} d(u^s) = d(u^*).$$

LOG-BARRIER FUNCTION AND INTERIOR-POINT METHODS

The $(\mathcal{ECO})$ problem is not combinatorial by nature because all constraints are active at the solution (by definition, active constraints are those that are satisfied as equalities). At the same time, the methods for $(\mathcal{ECO})$ require an initial approximation from the neighborhood $S(y^*, \delta)$. The $(\mathcal{ICO})$ in general and convex $(\mathcal{ICO})$ in particular, are combinatorial by nature because the set of active at the solution constraints is unknown *a priori*. On the other hand, for convex $(\mathcal{ICO})$ there is no need to have an initial approximation $y \in S(y^*, \delta)$.

Consider PD interior-point methods for convex optimization that are based on the classical log-barrier function $\beta : \operatorname{int}\Omega \times \mathbb{R}_+ \to \mathbb{R}$ defined by the formula

$$\beta(x, \mu) = f(x) - \mu \sum_{i=1}^{p} \ln c_i(x).$$

Let $\ln t = -\infty$ for $t \leq 0$; then for any given $\mu > 0$ the Frisch log-barrier function $F : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}$ is defined as follows:

$$F(x, \mu) = \begin{cases} \beta(x, \mu), & x \in \operatorname{int}\Omega; \\ \infty, & x \notin \operatorname{int}\Omega. \end{cases}$$

The classical sequential unconstrained minimization technique (SUMT) [5] finds the PD trajectory $\{y(\mu) = (x(\mu), u(\mu))\}$ by the following formulas:

$$\nabla_x F(x(\mu), \mu) = \nabla f(x(\mu)) - \sum_{i=1}^{p} \mu \left(c_i(x(\mu))\right)^{-1} \nabla c_i(x(\mu)) = 0 \quad (20)$$

and

$$u(\mu) = \left(u_i(\mu)) = \mu\left(c_i(x(\mu))\right)^{-1}, \; i = 1, \ldots, p\right). \quad (21)$$

Equation (20) is the optimality condition for the unconstrained problem

$$x(\mu) = \arg\min\left\{F(x, \mu) | x \in \mathbb{R}^n\right\}.$$

Thus, for a given $\mu > 0$, finding the PD approximation from Equations (20), (21) is equivalent to solving the following system for $(\hat{x}, \hat{u})$:

$$\nabla_x L(\hat{x}, \hat{u}) = \nabla f(\hat{x}) - \sum_{i=1}^{p} \hat{u}_i \nabla c_i(\hat{x}) = 0,$$
$$\hat{u}_i c_i(\hat{x}) = \mu, \quad i = 1, \dots, p. \tag{22}$$

It follows from Equation (20) that $x(\mu) \in \operatorname{int}\Omega$. It follows from Equation (21) that $u(\mu) \in \mathbb{R}^p_{++}$, that is $x(\mu)$ and $u(\mu)$ are primal and dual interior points. It follows from Equations (20), (21) that

$$\nabla_x F(x(\mu), \mu) = \nabla_x L(x(\mu), u(\mu)) = 0,$$

or

$$d(u(\mu)) = \min_{x \in \mathbb{R}^n} L(x, u(\mu)) = L(x(\mu), u(\mu))$$

and the PD gap (i.e., the difference between the values of the primal and dual objective functions) is

$$\begin{aligned}\Delta(\mu) &= f(x(\mu)) - d(u(\mu)) \\ &= \sum_{i=1}^{p} u_i(\mu) c_i(x(\mu)) = p\mu,\end{aligned}$$

that is

$$\lim_{\mu \to 0} x(\mu) = x^* \quad \text{and} \quad \lim_{\mu \to 0} u(\mu) = u^*. \tag{23}$$

Finding $(\hat{x}, \hat{u}) = (x(\mu), u(\mu))$ or its approximation from Equation (22) and reducing $\mu > 0$ is the main idea of the SUMT [5].

Solving the nonlinear PD system (22) is an infinite procedure in general. The main idea of the PD interior-point methods is to replace the nonlinear PD system (22) by one Newton step toward solution of the system (22) and follow that by the barrier parameter update. For a given approximation $y = (x, u)$ and the barrier parameter $\mu > 0$, the application of Newton's method to the nonlinear PD system (22) leads to the following linear PD system:

$$\begin{bmatrix} \nabla^2_{xx} L(x, u) & -\nabla c(x)^T \\ U \nabla c(x) & C(x) \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta u \end{bmatrix} = \begin{bmatrix} -\nabla L(x, u) \\ -Uc(x) + \mu e \end{bmatrix}, \tag{24}$$

where $C(x) = \operatorname{diag}\left(c_i(x)\right)_{i=1}^{p}$, $U = \operatorname{diag}\left(u_i\right)_{i=1}^{p}$, and $e = (1, \dots, 1)^T \in \mathbb{R}^p$.

The system (24) finds the Newton direction $\Delta y = (\Delta x, \Delta u)$, which is used to update the current approximation $y = (x, u)$:

$$\overline{x} = x + \alpha \Delta x; \quad \overline{u} = u + \alpha \Delta u. \tag{25}$$

One must determine the step length $\alpha > 0$ in such a way that a new approximation $\overline{y} = (\overline{x}, \overline{u})$ not only remains primal and dual interior, but also remains in the area for which Newton's method for PD system (22) is well defined for the updated barrier parameter [6]. For some classes of constrained optimization problems, including linear programming problems (LPs) and quadratic programming problems (QPs), it is possible to take $\alpha = 1$ in Equation (25) and update the barrier parameter μ by the formula $\overline{\mu} = \mu(1 - \rho/\sqrt{n})$, where $0 < \rho < 1$ is independent on n. The new approximation belongs to the neighborhood of the solution of the system (22) with μ replaced by $\overline{\mu}$. Moreover, each step reduces the PD gap by the same factor $(1 - \rho/\sqrt{n})$. This leads to polynomial complexity of the PD interior-point methods for LP and QP [6]. LP, QP, and QP with quadratic constraints problems are well structured, which means that constraints and epigraph of the objective function can be equipped with a self-concordant (SC) barrier [7,8].

If $(\mathcal{ICO})$ problem is not well structured, establishing polynomial complexity of the path-following methods becomes problematic, if not impossible. Nevertheless, the PD interior-point approach remains productive and leads to globally convergent and numerically efficient PD methods [9–13].

There are two main classes of PD methods for $(\mathcal{ICO})$: interior-point PD method based on the path-following idea (which goes back to SUMT) and exterior-point PD methods (which are based on nonlinear rescaling (NR) theory [14–16]).

PRIMAL–DUAL INTERIOR-POINT METHODS

Consider the following problem

$$\begin{aligned} &\min f(x), \\ &\text{s.t. } c(x) - w = 0, \\ &w \geq 0, \end{aligned} \tag{26}$$

which is equivalent to $(\mathcal{ICO})$ problem.

The log-barrier function $F(x, w, \mu) = f(x) - \mu \sum_{i=1}^{p} \ln w_i$ is used to handle the nonnegativity of the slack vector $w \in \mathbb{R}_+^p$, which guarantees the primal feasibility.

One step of the path-following method consists of solving the following $(\mathcal{ECO})$ problem:

$$\begin{aligned} &\min F(x, w, \mu), \\ &\text{s.t. } c(x) - w = 0, \end{aligned} \tag{27}$$

followed by the barrier parameter $\mu > 0$ update.

The Lagrangian for the problem (27) is defined by formula

$$L(x, w, u) = f(x) - \mu \sum_{i=1}^{p} \log w_i - \sum_{i=1}^{p} u_i (c_i(x) - w_i).$$

Let $W = \operatorname{diag}(w_i)_{i=1}^p$ and $U = \operatorname{diag}(u_i)_{i=1}^p$. The following Lagrange system of equations corresponds to $(\mathcal{ECO})$ (27).

$$\begin{aligned} \nabla f(x) - \nabla c(x)^T u &= 0 \\ -\mu e + WUe &= 0 \\ c(x) - w &= 0 \end{aligned} \tag{28}$$

Application of Newton's method to nonlinear PD system (28) leads to the following linear PD system for finding the Newton directions:

$$\begin{bmatrix} \nabla_{xx}^2 L(x, u) & 0 & -\nabla c(x)^T \\ 0 & U & W \\ \nabla c(x) & -I_p & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta w \\ \Delta u \end{bmatrix} = \begin{bmatrix} -\nabla f(x) + \nabla c(x)^T u \\ \mu e - WUe \\ -c(x) + w \end{bmatrix}, \tag{29}$$

where $\nabla_{xx}^2 L(x, u) = \nabla^2 f(x) - \sum_{i=1}^{p} u_i \nabla^2 c_i(x)$ is the Hessian in x of the Lagrangian for $(\mathcal{ICO})$ problem.

For convex $(\mathcal{ICO})$ the merit function

$$\nu(y) \equiv \nu(x, w, u) = \max\left\{ \|\nabla_x L(x, w, u)\|, \|c(x) - w\|, \|WUe\|, \right\}$$

is used. Under the standard second-order optimality conditions, the merit function $\nu(y)$ satisfies Equations (7) and (8).

Take $\delta > 0$ to be sufficiently small and $0 < \epsilon \ll \delta$ to be the defined accuracy. Consider the following PD interior-point method:

PDIPM - Primal–Dual Interior-Point Method

Step 0. Let $y \in S(y^*, \delta) = \{y : \|y - y^*\| \leq \delta\}$ be the initial approximation.

Step 1. If $\nu(x, w, u) \leq \epsilon$, Output $y = (x, u)$ as a solution.

Step 2. Calculate the barrier parameter $\mu = \min\left\{\theta\mu,\ \nu^2(x, w, u)\right\}$, $0 < \theta < 1$.

Step 3. Find Δx, Δw, and Δu from (29).

Step 4. Calculate the parameter κ and the step lengths α_P and α_D by formulas

$$\kappa = \max\left\{\bar{\kappa}, 1 - \nu(x, w, u)\right\}, \quad 0 < \bar{\kappa} < 1,$$

$$\alpha_P = \min_{1 \leq i \leq m} \left\{ 1;\ -\kappa \frac{(w^s)_i}{(\Delta w^s)_i} : (\Delta w^s)_i < 0 \right\},$$

$$\alpha_D = \min_{1 \leq i \leq m} \left\{ 1;\ -\kappa \frac{(u^s)_i}{(\Delta u^s)_i} : (\Delta u^s)_i < 0 \right\}.$$

Step 5. Update the primal–dual pair by the formulas

$$\hat{x} := x + \alpha_P \Delta x, \quad \hat{w} := w + \alpha_P \Delta w, \quad \hat{u} := u + \alpha_D \Delta u.$$

Step 6. Goto Step 1.

If the standard second-order optimality conditions are satisfied for $(\mathcal{ICO})$, the Hessians $\nabla^2 f(x)$ and $\nabla^2 c_i(x)$, $i = 1, \ldots, p$ are Lipschitz continuous, $\delta > 0$ is sufficiently small, and a starting point $y \in S(y^*, \delta)$, then the PD sequence generated by the PDIPM converges to the KKT's point with a quadratic rate [10,17].

PRIMAL–DUAL NONLINEAR RESCALING METHOD

Consider a class Ψ of strictly concave and twice continuously differential functions $\psi : (t_0, t_1) \to \mathbb{R}$, $-\infty < t_0 < 0 < t_1 < \infty$ that satisfy the following properties:

$1^0. \psi(0) = 0.$

$2^0. \psi'(t) > 0.$

$3^0. \psi'(0) = 1.$

$4^0. \psi''(t) < 0.$

$5^0.$there *is* $a > 0$ that $\psi(t) \leq -at^2$, $t \leq 0$.

$6^0. a) \psi'(t) \leq b_1 t^{-1}, b) - \psi''(t) \leq b_2 t^{-2}, t > 0,$

$b_1 > 0, b_2 > 0.$

The following transformations $\psi \in \Psi$ satisfy the above properties:

1. Exponential transformation [18]

$$\psi_1(t) = 1 - e^{-t}.$$

2. Logarithmic MBF [14]

$$\psi_2(t) = \ln(t + 1).$$

3. Hyperbolic MBF [14]

$$\psi_3(t) = \frac{t}{1 + t}.$$

Each of the above transformations can be modified in the following way. For a given $-1 < \tau < 0$ we define quadratic extrapolation of the transformations $1 - 3$ by the formulas

4.

$$\psi_{q_i}(t) = \begin{cases} \psi_i(t), & t \geq \tau, \\ q_i(t) = a_i t^2 + b_i t + c_i, & t \leq \tau, \end{cases}$$

where a_i, b_i, c_i one finds from the following equations: $\psi_i(\tau) = q_i(\tau)$, $\psi_i'(\tau) = q_i'(\tau)$, $\psi_i''(\tau) = q_i''(\tau)$.

Modification 4 leads to transformations that are defined on $(-\infty, \infty)$ and, along with penalty function properties, have some additional important features [15,16,19].

Due to $1^0 - 3^0$ for any $\psi \in \Psi$ and any $k > 0$, the following problem

$$f(x^*) = \min \Big\{ f(x) | k^{-1} \psi(kc_i(x)) \geq 0, \\ i = 1, \ldots, p \Big\} \quad (30)$$

is equivalent to the original $(\mathcal{ICO})$ problem. The classical Lagrangian $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^p_+ \times \mathbb{R}_+ \to \mathbb{R}^1$ for the equivalent problem (30) is defined by the formula

$$\mathcal{L}(x, u, k) = f(x) - k^{-1} \sum_{i=1}^{p} u_i \psi(kc_i(x)).$$

The NR principle consists of transforming the original $(\mathcal{ICO})$ problem into an equivalent one and then using the classical Lagrangian for the equivalent problem for theoretical analysis and numerical methods. In contrast to SUMT, the NR principle leads to exterior-point methods. Convergence of the NR methods is due to both, the Lagrange multipliers and the scaling parameter updates.

We use the following merit function:

$$\nu \equiv \nu(x,u) = \max\left\{ \|\nabla_x L(x,u)\|, \, -\min_{1\le i\le p} c_i(x), \right.$$
$$\left. \sum_{i=1}^{p} |u_i||c_i(x)|, \, -\min_{1\le i\le p} u_i, \right\}. \quad (31)$$

For convex $(\mathcal{ICO})$ under the standard second-order optimality condition, the merit function (31) satisfies Equations (7) and (8).

For a given $u \in \mathbb{R}^p_{++}$ and $k > 0$, one step of the NR method with scaling parameter update consists in finding

$$\hat{x} = \arg\min \left\{ \mathcal{L}(x,u,k) \,|\, x \in \mathbb{R}^n \right\}, \quad (32)$$

or

$$\hat{x} : \nabla_x \mathcal{L}(\hat{x},u,k) = \nabla f(\hat{x}) - \sum_{i=1}^{p} \psi'\left(kc_i(\hat{x})\right)$$
$$u_i \nabla c_i(\hat{x}) = 0,$$

followed by the Lagrange multipliers update by formula

$$\hat{u}_i = \psi'\left(kc_i(\hat{x})\right) u_i, i = 1, \ldots, p,$$

or

$$\hat{u} = \Psi'\left(kc(\hat{x})\right) u, \quad (33)$$

where $\Psi'\left(k_s c(x)\right) = diag\left(\psi'\left(kc_i(x)\right)\right)_{i=1}^{p}$, and scaling parameter update by formula

$$\hat{k} := \nu(\hat{x},\hat{u})^{-0.5}. \quad (34)$$

In other words, for a given Lagrange multipliers vector $u \in \mathbb{R}^p_{++}$ and the scaling parameter $k > 0$, one step of the NR method is equivalent to solving the following nonlinear PD system:

$$\nabla_x \mathcal{L}(\hat{x},u,k) = \nabla f(\hat{x}) - \sum_{i=1}^{p} \psi'\left(kc_i(\hat{x})\right) u_i \nabla c_i(\hat{x})$$
$$= \nabla_x L(\hat{x},\hat{u}) = 0, \quad (35)$$
$$\hat{u} = \Psi'\left(kc(\hat{x})\right) u, \quad (36)$$

for $\hat{x}$ and $\hat{u}$, followed by the scaling parameter $k > 0$ update by Equation (34).

Application of Newton's method for solving the nonlinear PD system (35)–(36) for $\hat{x}$ and $\hat{u}$ leads to the primal–dual nonlinear rescaling (PDNR) method for solving $(\mathcal{ICO})$ problem [20].

By linearizing Equations (35)–(36) and assuming that $\overline{u} = \Psi'\left(kc(x)\right) u$, we obtain the following linear PD system for finding the Newton direction $(\Delta x, \Delta u)$:

$$\begin{bmatrix} \nabla^2_{xx} L(\cdot) & -\nabla c^T(\cdot) \\ -U\Psi''(\cdot)\nabla c(\cdot) & \frac{1}{k} I_p \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta u \end{bmatrix} = \begin{bmatrix} -\nabla_x L(\cdot) \\ \frac{1}{k}(\overline{u} - u) \end{bmatrix}, \quad (37)$$

where $\nabla c(\cdot) = \nabla c(x)$, $\Psi'(\cdot) = \Psi''\left(kc(x)\right) = diag\left(\psi''\left(kc_i(x)\right)\right)_{i=1}^{p}$, $U = diag\,(u_i)_{i=1}^{p}$ and I_p is an identity matrix in $\mathbb{R}^p$.

Let $\delta > 0$ be small enough and $0 < \epsilon \ll \delta$ be the desired accuracy. Then for a given $x \in \mathbb{R}^n$, Lagrange multipliers vector $u \in \mathbb{R}^p_{++}$, and scaling parameter $k > 0$, one step of the PDNR method consists of the following operations:

PDNRM - Primal–Dual Nonlinear Rescaling Method

Step 0. Let $y = (x,u) \in S(y^*, \delta)$.
Step 1. If $\nu(y) \le \epsilon$, then output y as a solution.
Step 2. Set $k = \nu(x,u)^{-0.5}$.
Step 3. Find Δx and Δv from (37).
Step 4. Update the primal–dual pair by the formulas

$$\hat{x} := x + \Delta x, \quad \hat{u} := u + \Delta u.$$

Step 5. Goto Step 1.

If the standard second-order optimality conditions are satisfied, the Lipschitz conditions for the Hessians $\nabla^2 f(x)$, $\nabla^2 c_i$, $i = 1, \ldots, p$ hold, and $y = (x, u) \in S(y^*, \delta)$, then the PDNR algorithm generates a PD sequence that converges to the PD solution with 1.5-Q-superlinear rate. A small modification of the PDNR algorithm generates a PD sequence that converges from any starting point $y = (x, u) \in \mathbb{R}^n \times \mathbb{R}^p_{++}$ with asymptotic 1.5-Q-superlinear rate [20].

PRIMAL–DUAL EXTERIOR-POINT METHOD FOR INEQUALITY CONSTRAINED OPTIMIZATION

Due to properties 1^0 and 2^0 for any given vector $k = (k_1, \ldots, k_p) \in \mathbb{R}^p_{++}$, we have

$$c_i(x) \geq 0 \Longleftrightarrow k_i\psi(k_i c_i(x)) \geq 0, \quad i = 1, \ldots, p.$$

Consider the Lagrangian $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^p_+ \times \mathbb{R}^p_{++} \to \mathbb{R}$

$$\mathcal{L}(x, u, \mathbf{k}) = f(x) - \sum_{i=1}^{p} k_i^{-1} u_i \psi(k c_i(x))$$

for the equivalent problem. Let $k > 0, x \in \mathbb{R}^n$, $u \in \mathbb{R}^p$ and $\mathbf{k} = (k_i = k(u_i)^{-1}, i = 1, \ldots, p)$.

One step of the NR method with "dynamic" scaling parameters update maps the given triple $(x, u, \mathbf{k}) \in \mathbb{R}^n \times \mathbb{R}^p_+ \times \mathbb{R}^p_{++}$ into the triple $(\hat{x}, \hat{u}, \hat{\mathbf{k}}) \in \mathbb{R}^n \times \mathbb{R}^p_+ \times \mathbb{R}^p_{++}$ defined by formulas

$$\begin{aligned} \hat{x} : \nabla_x \mathcal{L}(\hat{x}, u, \mathbf{k}) &= \nabla f(\hat{x}) - \sum_{i=1}^{p} \psi'(k_i c_i(\hat{x})) u_i \nabla c_i(\hat{x}) \\ &= \nabla f(\hat{x}) - \sum_{i=1}^{p} \hat{u}_i \nabla c_i(\hat{x}), \end{aligned} \tag{38}$$

$$\hat{u} : \hat{u}_i = u_i \psi'(k_i c_i(\hat{x})), \quad i = 1, \ldots, p, \tag{39}$$

$$\hat{\mathbf{k}} : \hat{k}_i = k\hat{u}_i^{-1}, \quad i = 1, \ldots, p. \tag{40}$$

Such a method was considered in Ref. 18 for exponential transformation ψ_1. By removing the formula for the scaling vector update (40) from the system (38), (39), (40) we obtain the following nonlinear PD system:

$$\nabla_x L(\hat{x}, \hat{u}) = \nabla f(\hat{x}) - \sum_{i=1}^{p} \hat{u}_i \nabla c_i(\hat{x}), \tag{41}$$

$$\hat{u} = \Psi'(kc(\hat{x}))u, \tag{42}$$

where $\Psi'(kc(\hat{x})) = \text{diag}(\psi'(k_i c_i(\hat{x})))_{i=1}^p$. To measure the distance between the current approximation $y = (x, u)$ and the solution y^*, the merit function (31) is used.

Due to Equation (40), the Lagrange multipliers that correspond to the passive constraints (LMPC) converge to zero with at least a quadratic rate. Therefore, for $0 < \epsilon \ll 1$, it requires at most $s = \mathcal{O}(\ln \ln \epsilon^{-1})$ Lagrange multipliers updates for the LMPC to become of the order of $\mathcal{O}(\epsilon^2)$. So the PD system (41), (42) is reduced to the following nonlinear PD system:

$$\nabla_x L(\hat{x}, \hat{u}) = \nabla f(\hat{x}) - \sum_{i=1}^{r} \hat{u}_i \nabla c_i(\hat{x}), \tag{43}$$

$$\hat{u}_i = u_i \psi'(k_i c_i(\hat{x})), \quad i = 1, \ldots, r, \tag{44}$$

where $I^* = \{i : c_i(x^*) = 0\} = \{1, \ldots, r\}$.

Application of Newton's method for solving the PD system (43), (44) leads to a local PDEP method for $(\mathcal{ICO})$ problems. Start with linearization of the system (44); due to the property 2^0 of the transformation $\psi \in \Psi$, the inverse ψ'^{-1} exists. Therefore using the identity $\psi'^{-1} = \psi^{*\prime}$, where $\psi^*(s) = \inf_t\{st - \psi(t) | t \in \mathbb{R}\}$ is the conjugate of ψ, and denoting $\varphi = -\psi^*$, we can rewrite Equation (44) as follows:

$$\begin{aligned} c_i(\hat{x}) &= k^{-1} u_i \psi'^{-1}\left(\frac{\hat{u}_i}{u_i}\right) = k^{-1} u_i \psi^{*\prime}\left(\frac{\hat{u}_i}{u_i}\right) \\ &= -k^{-1} u_i \varphi'\left(\frac{\hat{u}_i}{u_i}\right). \end{aligned}$$

It follows from property 3^0 of transformation ψ that $\varphi'(1) = 0$. Assuming $\hat{x} = x + \Delta x$ and $\hat{u} = u + \Delta u$, and ignoring terms of the second and higher order, we obtain

$$c_i(\hat{x}) = c_i(x) + \nabla c_i(x) \Delta x$$

$$= -k^{-1}u_i\varphi'\left(\frac{u_i+\Delta u_i}{u_i}\right)$$
$$= -k^{-1}u_i\varphi'\left(1+\frac{\Delta u_i}{u_i}\right) = -k^{-1}\varphi''(1)\Delta u_i,$$
$$i = 1,\ldots,r,$$

or

$$c_i(x) + \nabla c_i(x)\Delta x + k^{-1}\varphi''(1)\Delta u_i = 0,$$
$$i = 1,\ldots,r.$$

By linearizing the system (43) at $y = (x, u)$, we obtain the following linear PD system for finding the PD Newton directions:

$$\begin{bmatrix} \nabla_{xx}L(\cdot) & -\nabla c^T(\cdot) \\ \nabla c(\cdot) & k^{-1}\varphi''(1)I_r \end{bmatrix}\begin{bmatrix} \Delta x \\ \Delta u \end{bmatrix} = \begin{bmatrix} \nabla_x L(\cdot) \\ -c(\cdot) \end{bmatrix} \quad (45)$$

where I_r is the identity matrix in $\mathbb{R}^r$ and $\nabla c(x) = J(c(x))$ is the Jacobian of the vector function $c(x)$.

Let $\delta > 0$ be small enough and $0 < \epsilon \ll \delta$ be the desired accuracy. Then the PDEP method for $(\mathcal{ICO})$ problems consists of the following operations.

PDEPICOM - Primal–Dual Exterior-Point Method for $\mathcal{ICO}$

Step 0. Let $y = (x, u) \in S(y^*, \delta)$.
Step 1. If $\nu(y) \leq \epsilon$, then output y as a solution.
Step 2. Set $k = \nu(x, u)^{-1}$.
Step 3. Find Δx and Δv from (45).
Step 4. Update the primal–dual pair by the formulas

$$\hat{x} := x + \Delta x, \quad \hat{u} := u + \Delta u.$$

Step 5. Goto Step 1.

Under the standard second-order optimality conditions and Lipschitz conditions for the Hessians $\nabla^2 f(x)$, $\nabla^2 c_i$, $i = 1,\ldots,r$, the PDEP algorithm generates a PD sequence that converges to the PD solution $y^* = (x^*, u^*)$ with quadratic rate. The globally convergent PDEP method with an asymptotic quadratic rate is given in [21].

PRIMAL–DUAL EXTERIOR-POINT METHOD FOR OPTIMIZATION PROBLEMS WITH BOTH EQUALITY AND INEQUALITY CONSTRAINTS

Consider $p + q + 1$ twice continuously differential functions f, c_i, g_j : $\mathbb{R}^n \to \mathbb{R}$, $i = 1,\ldots,p, j = 1,\ldots,q$ and the feasible set

$$\Omega = \big\{x : c_i(x) \geq 0,\, i = 1,\ldots,p;\; g_j(x) = 0, j = 1,\ldots,q\big\}.$$

The problem with both inequality and equality constraints consists of finding

$$(\mathcal{IECO}) \qquad f(x^*) = \min\{f(x)|x \in \Omega\}.$$

We use $\psi \in \Psi$ to transform the inequality constraints $c_i(x) \geq 0$, $i = 1,\ldots,p$ into an equivalent set of constraints. For any fixed $k > 0$, the following problem is equivalent to the original $(\mathcal{IECO})$ problem due to the properties of $\psi \in \Psi$; that is,

$$f(x^*) = \min\{f(x)|k^{-1}\psi(kc_i(x)) \geq 0,\; i = 1,\ldots,p;\; g_j(x) = 0, j = 1,\ldots,q\}.$$

For a given $k > 0$, the AL $\mathcal{L}_k : \mathbb{R}^n \times \mathbb{R}^p_+ \times \mathbb{R}^q \to \mathbb{R}^1$ for the equivalent problem is defined by the formula

$$\mathcal{L}_k(x, u, v) = f(x) - k^{-1}\sum_{i=1}^{p} u_i\psi(kc_i(x)) - \sum_{j=1}^{q} v_j g_j(x) + \frac{k}{2}\sum_{j=1}^{q} g_j^2(x). \quad (46)$$

The first two terms define the classical Lagrangian for the equivalent problem in the absence of equality constraints. The last two terms coincide with the AL terms associated with equality constraints.

For a given $k > 0$ and $y = (x, u, v) \in S(y^*, \delta)$, a step of nonlinear rescaling-AL method maps the triple $y = (x, u, v)$ into the triple $\hat{y} = (\hat{x}, \hat{u}, \hat{v})$, where

$$\hat{x} = \arg \min_{x \in \mathbb{R}^n} \mathcal{L}_k(x, u, v), \tag{47}$$

or

$$\hat{x} : \nabla f(\hat{x}) - \sum_{i=1}^{p} u_i \psi' \left(kc_i(\hat{x})\right) \nabla c_i(\hat{x})$$

$$- \sum_{j=1}^{q} \left(v_j - kg_j(\hat{x})\right) \nabla g_j(\hat{x}) = 0, \tag{48}$$

$$\hat{u}_i = u_i \psi'(kc_i(\hat{x})), \quad i = 1, \ldots, p; \tag{49}$$

$$\hat{v}_j = v_j - kg_j(\hat{x}), \quad j = 1, \ldots, q. \tag{50}$$

The system (48)–(50) can be replaced by the following nonlinear PD system:

$$\nabla_x L(\hat{x}, \hat{u}, \hat{v}) = \nabla f(\hat{x}) - \sum_{i=1}^{p} \hat{u}_i \nabla c_i(\hat{x})$$

$$- \sum_{j=1}^{q} \hat{v}_j \nabla g_j(\hat{x}) = 0, \tag{51}$$

$$\hat{u} - \Psi' \left(kc(\hat{x})\right) u = 0, \tag{52}$$

$$\hat{v} - v + kg(\hat{x}) = 0, \tag{53}$$

where $\Psi'(kc(\hat{x})) = \mathrm{diag}(\psi'(kc_i(\hat{x})))_{i=1}^{p}$.

We use the following merit function:

$$\nu(y) = \nu(x, u, v)$$

$$= \max\{\|\nabla_x L(x, u, v)\|, \quad -\min_{1 \le i \le p} c_i(x),$$

$$\max_{1 \le i \le q} |g_i(x)|, \quad \sum_{i=1}^{p} |u_i||c_i(x)|, \quad -\min_{1 \le i \le p} u_i\}. \tag{54}$$

Under the standard second-order optimality conditions, $\nu(y)$ possesses properties (7), (8).

Application of Newton's method to the system (51), (52), (53) for $\hat{x}$, $\hat{u}$, and $\hat{v}$ from the starting point $y = (x, u, v) \in S(y^*, \delta)$ leads to the PD exterior-point method (PDEPM) [22]. By linearizing the system (51), (52), (53) and ignoring terms of the second and higher order, we obtain the following linear PD system for finding the Newton direction $\Delta y = (\Delta x, \Delta u, \Delta v)$:

$$\begin{bmatrix} \nabla^2_{xx} L(\cdot) & -\nabla c^T(\cdot) & -\nabla g^T(\cdot) \\ -U\Psi''(\cdot)\nabla c(\cdot) & \frac{1}{k} I_p & 0 \\ \nabla g(\cdot) & 0 & \frac{1}{k} I_q \end{bmatrix}$$

$$\times \begin{bmatrix} \Delta x \\ \Delta u \\ \Delta v \end{bmatrix} = \begin{bmatrix} -\nabla_x L(\cdot) \\ \frac{1}{k}(\overline{u} - u) \\ -g(\cdot) \end{bmatrix}, \tag{55}$$

where $\nabla c(\cdot) = \nabla c(x)$, $\nabla g(\cdot) = \nabla g(x)$, $\Psi'(\cdot) = \Psi''\big(kc(x)\big) = \mathrm{diag}\,\big(\psi''\big(kc_i(x)\big)\big)_{i=1}^{p}$, $\overline{u} = \Psi'\big(kc(x)\big)u$, $U = \mathrm{diag}\,(u_i)_{i=1}^{p}$, and I_p, I_q are the identity matrices in $\mathbb{R}^p$ and $\mathbb{R}^q$ respectively.

The PD exterior-point method for constrained optimization problems with both inequality and equality constraints consists of the following operations.

PDEPM - Primal–Dual Exterior-Point Method

Step 0. Let $y \in S(y^*, \delta)$..

Step 1. If $\mu(y) \le \epsilon$, then output y as a solution.

Step 2. Set $k = \nu(x, u, v)^{-0.5}$.

Step 3. Find Δx, Δu, and Δv from Equation (55).

Step 4. Update the primal–dual pair by the formulas

$$\hat{x} := x + \Delta x, \quad \hat{u} := u + \Delta u. \quad \hat{v} := v + \Delta v.$$

Step 5. Goto Step 1.

Under the standard second-order optimality conditions and Lipschitz conditions for the Hessians $\nabla^2 f(x)$, $\nabla^2 c_i$, $i = 1, \ldots, m$, $\nabla^2 g_i$, $i = 1, \ldots, q$, for any $y \in S(y^*, \delta)$, the PDEPM algorithm generates the PD sequence that converges to the PD solution with 1.5-Q-superlinear rate [22].

CONCLUDING REMARKS

As we have seen, any PD method for constrained optimization is associated with particular PD nonlinear system of equations. Application of Newton's method to the system is equivalent to one step of either an interior- or an exterior-point method. Therefore, the computational process is defined by a particular PD nonlinear system and excellent convergence properties of Newton's method. As a result, under the standard second-order optimality conditions, the PD methods generate sequences that converge to the PD solution with an asymptotic quadratic rate.

As a practical matter it is worth mentioning that in the neighborhood of the solution the PD interior-point method generates a sequence similar to that generated by Newton's method for solving KKT's system, while the PD exterior-point method generates a sequence similar to that generated by Newton's method for solving the Lagrange system of equations that correspond to the active constraints. Unlike the former nonlinear system, the latter system does not contain complementarity equations. Therefore, the practical PD exterior-point method often is capable of finding the PD solution with a very high of accuracy.

REFERENCES

1. Polyak BT. Introduction to optimization. New York: Software Inc.; 1987.
2. Hestenes MR. Multipliers and gradient methods. J Optim Theory Appl 1969;4(5):303–320.
3. Powell MJD. A method for nonlinear constraints in minimization problems. In: Fletcher R, editor. Optimization. London: Acadamic Press; 1969. pp. 283–298.
4. Polyak RA. On the local quadratic convergence of the primal–dual augmented Lagrangian method. Optim Method Soft 2009;24(3):369–379.
5. Fiacco AV, McCormick GP. Nonlinear programming: sequential unconstrained minimization techniques. New York: John Wiley & Sons, Inc.; 1968.
6. Wright SJ. Primal–dual interior-point methods. Philadelphia (PA): SIAM; 1997.
7. Nesterov Y, Nemirovsky A. Interior-point polynomial algorithm in convex programming. Philadelphia (PA): SIAM; 1994.
8. Nesterov Y. Introductory lectures on convex optimization. Norwell (MA): Kluwer Academic Publishers; 2004.
9. Armand P, Benoist J, Orban D. Dynamic updates of the barrier parameter in primal–dual methods for nonlinear programming. Comput Optim Appl 2008;41(1): 1–25.
10. Griva I, Shanno DF, Vanderbei RJ, *et al*. Global convergence of a primal–dual interior-point method for nonlinear programming. Algorithm Oper Res 2008;3(1):12–29.
11. Forsgren A, Gill P. Primal–dual interior methods for nonconvex nonlinear programming. SIAM J Optim 1998;8(4):1132–1152.
12. Wachter A, Biegler LT. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Math Program 2006;106(1):25–57.
13. Waltz RA, Morales JL, Nocedal J, *et al*. An interior algorithm for nonlinear optimization that combines line search and trust region steps. Math Program 2006;107:391–408.
14. Polyak RA. Modified barrier functions theory and methods. Math Program 1992;54:177–222.
15. Polyak R, Teboulle M. Nonlinear rescaling and proximal-like methods in convex optimization. Math Program 1997;76:265–284.
16. Polyak RA. Nonlinear rescaling vs. smoothing technique in convex optimization. Math Program 2002;92:197–235.
17. Yamashita H, Yabe H. Superlinear and quadratic convergence of some primal–dual interior point methods for constrained optimization. Math Program 1996;75:377–397.
18. Tseng P, Bertsekas D. On the convergence of exponential multipliers method for convex programming. Math Program 1993;60:1–19.
19. Ben-Tal A, Zibulevsky M. Penalty-barrier methods for convex optimization. SIAM J Control Optim 1997;7:347–366.

20. Griva I, Polyak RA. Primal–dual nonlinear rescaling method with dynamic scaling parameter update. Math Program 2006;106(2):237–259.

21. Polyak RA. Primal–dual exterior-point method for convex optimization. Optim Method Softw 2008;23(1):141–160.

22. Griva I, Polyak RA. 1.5-Q-superlinear convergence of an exterior-point method for constrained optimization. J Glob Optim 2008;40(4):679–695.