# Approximate Subgradient Methods for Lagrangian Relaxations on Networks

Eugenio Mijangos

**Abstract** Nonlinear network flow problems with linear/nonlinear side constraints can be solved by means of Lagrangian relaxations. The dual problem is the maximization of a dual function whose value is estimated by minimizing approximately a Lagrangian function on the set defined by the network constraints. We study alternative stepsizes in the approximate subgradient methods to solve the dual problem. Some basic convergence results are put forward. Moreover, we compare the quality of the computed solutions and the efficiency of these methods.

## 1 Introduction

Consider the nonlinearly constrained network flow problem **(NCNFP)**

$$\underset{x}{\text{minimize}} \;\; f(x) \tag{1}$$
$$\text{subject to } \; x \in \mathscr{F} \tag{2}$$
$$c(x) \leq 0, \tag{3}$$

where:

- $\mathscr{F} = \{x \in \mathbf{R}^n \mid Ax = b,\ 0 \leq x \leq \bar{x}\}$, where $A$ is a node-arc incidence $m \times n$-matrix, $b$ is the production/demand $m$-vector, $x$ are the flows on the arcs of the network represented by $A$, and $\bar{x}$ is the vector of capacity bounds imposed on the flow of each arc.
- The side constraints (3) are defined by $c : \mathbf{R}^n \rightarrow \mathbf{R}^r$, such that $c = [c_1, \cdots, c_r]^T$, where $c_i(x)$ is linear or nonlinear and twice continuously differentiable on $\mathscr{F}$ for all $i = 1, \cdots, r$.

Eugenio Mijangos
University of the Basque Country, Department of Applied Mathematics, Statistics and Operations Research, Spain, e-mail: eugenio.mijangos@ehu.es

- $f : \mathbf{R}^n \to \mathbf{R}$ is nonlinear and twice continuously differentiable on $\mathscr{F}$.

We focus on the primal problem **NCNFP** and its dual problem

$$\text{maximize} \quad q(\mu) = \min_{x \in \mathscr{F}} l(x, \mu) \tag{4}$$

$$\text{subject to} \quad \mu \in \mathscr{M}, \tag{5}$$

where the Lagrangian function is

$$l(x, \mu) = f(x) + \mu^T c(x) \tag{6}$$

and $\mathscr{M} = \{\mu \mid \mu \geq 0,\ q(\mu) > -\infty\}$. We assume throughout this paper that the constraint set $\mathscr{M}$ is closed and convex. Since $q$ is concave on $\mathscr{M}$, it is continuous on $\mathscr{M}$. When exact values of $q$ are used, we assume that for every $\mu \in \mathscr{M}$ some vector $x(\mu)$ that minimizes $l(x, \mu)$ over $x \in \mathscr{F}$ can be calculated, yielding a subgradient $c(x(\mu))$ of $q$ at $\mu$, which allows to solve **NCNFP** by using primal-dual methods, see [2]. Nevertheless, a substantial drawback of this kind of methods is the need to obtain at each iteration an exact solution to the subproblem included in (4). In this paper in order to allow for inexact solution of this minimization, we consider *approximate subgradient methods* [6, 8, 7] in the solution of this problem. The basic difference between these methods and the classical subgradient methods is that they replace the subgradients with inexact subgradients.

Given a scalar $\varepsilon \geq 0$ and a vector $\overline{\mu} \in \mathscr{M}$, we say that $c$ is an $\varepsilon$-*subgradient* (approximate subgradient) at $\overline{\mu}$ if

$$q(\mu) \leq q(\overline{\mu}) + \varepsilon + c^T(\mu - \overline{\mu}), \qquad \forall \mu \in \mathbf{R}^r. \tag{7}$$

The set of all $\varepsilon$-subgradients at $\overline{\mu}$ is the $\varepsilon$-*subdifferential* at $\overline{\mu}$ (i.e. $\partial_\varepsilon q(\overline{\mu})$).

In our context, we minimize approximately $l(x, \mu^k)$ over $x \in \mathscr{F}$ by efficient techniques specialized for networks [15], obtaining a vector $x^k \in \mathscr{F}$ with

$$l(x^k, \mu^k) \leq \inf_{x \in \mathscr{F}} l(x, \mu^k) + \varepsilon_k. \tag{8}$$

As is shown in [2, 8], the corresponding constraint vector, $c(x^k)$, is an $\varepsilon_k$-subgradient at $\mu^k$. If we denote $q_{\varepsilon_k}(\mu^k) = l(x^k, \mu^k)$, by definition of $q(\mu^k)$ and using (8) we have

$$q(\mu^k) \leq q_{\varepsilon_k}(\mu^k) \leq q(\mu^k) + \varepsilon_k \qquad \forall k. \tag{9}$$

An approximate subgradient method is defined by

$$\mu^{k+1} = [\mu^k + \alpha_k c^k]^+, \tag{10}$$

where $c^k$ is an approximate subgradient at $\mu^k$, $[\cdot]^+$ denotes the projection on the closed convex set $\mathscr{M}$, and $\alpha_k$ is a positive scalar stepsize.

Different ways of computing the stepsize have been considered:

(a) Constant step rule (CSR) with Shor-type scaling [14].
(b) A variant of the constant step rule (VCSR) of Shor.
(c) Diminishing stepsize rule with scaling (DSRS) [13, 5, 14].
(d) The diminishing stepsize rule without scaling (DSR) suggested by Correa and Lemaréchal in [3].
(e) A dynamically chosen stepsize rule based on an estimation of the optimal value of the dual function by means of an adjustment procedure (DSAP) similar to that suggested by Nedić and Bertsekas in [12] for incremental subgradient methods.

The convergence of these methods was studied in the cited papers for the case of exact subgradients. The convergence of the approximate subgradient methods was analyzed by Kiwiel [6].

An alternative study of the convergence of some of these methods and their application in the solution of nonlinear networks was carried out in [8, 7].

In this work some basic convergence results obtained by Shor [14] are extended to approximate subgradient methods. Moreover, we compare the quality of the computed solution and the efficiency of the approximate subgradient methods when using CSR, VCSR, DSRS, DSR, and DSAP over **NCNFP** problems.

This paper is organized as follows: Sect. 2 presents the stepsize rules with the corresponding convergence results; Sect. 3, the solution to the nonlinearly constrained network flow problem; Sect. 4 puts forward the numerical tests; and Sect. 4 displays the conclusions.

## 2 Stepsize Rules and Convergence Results

Throughout this section, we use the notation

$$q^* = \sup_{\mu \in \mathscr{M}} q(\mu), \quad \mathscr{M}^* = \{\mu \in \mathscr{M} \mid q(\mu) = q^*\}, \tag{11}$$

and $\|\cdot\|$ denotes the standard Euclidean norm.

**Assumption 1** (subgradient boundedness). There exists a scalar $C > 0$ such that for $\mu^k \in \mathscr{M}$, $\varepsilon_k \geq 0$ and $c^k \in \partial_{\varepsilon_k} q(\mu^k)$, we have $\|c^k\| \leq C$, for $k = 0, 1, \ldots$.

We say that $\overline{\mu}$ is an $\varepsilon$-*optimal solution* of the dual problem when $0 \in \partial_\varepsilon q(\overline{\mu})$, i.e. when $q(\overline{\mu}) \geq q^* - \varepsilon$.

In this paper various kinds of stepsize rules have been considered.

### *2.1 Constant Step Rule (CSR)*

As is well known the classical scaling of Shor (see [14])

$$\alpha_k = \frac{s_k}{\|c^k\|} \tag{12}$$

with $s_k = s$ gives rise to an $s$-constant-step algorithm.

Note that constant stepsizes (i.e. $\alpha_k = s$ for all $k$) are unsuitable because the function $q$ may be nondifferentiable at the optimal point and then $\{c^k\}$ does not necessarily tend to zero, even if $\{\mu^k\}$ converges to the optimal point, see [14].

Next, we show some basic convergence results when $c^k$ is an approximate subgradient, which are similar to the results obtained by Shor [14] in the case of exact subgradients.

**Proposition 1.** *Consider the $\varepsilon$-subgradient iteration*

$$\mu^{k+1} = \left[\mu^k + \alpha_k c^k\right]^+, \tag{13}$$

*where $\alpha_k = s_k/\|c^k\|$ and $s_k = s > 0$ for any $k$, and $c^k \in \partial q_{\varepsilon_k}(\mu^k)$, with $\lim_{k\to\infty} \varepsilon_k = \varepsilon \geq 0$. Then, for any $\delta > 0$ and any dual optimal solution $\mu^* \in \mathscr{M}^*$, either one can find $k = \bar{k}$, where $\mu^{\bar{k}}$ is an $\varepsilon_k$-optimal solution, or there exist an index $\bar{k}$ and a point $\bar{\mu} \in \mathscr{M}$ such that $q(\bar{\mu}) = q(\mu^{\bar{k}}) + \varepsilon_{\bar{k}}$ and $\|\bar{\mu} - \mu^*\| < \frac{s}{2}(1+\delta)$.*

*Proof.* Let $\mu^* \in \mathscr{M}^*$ and let $\delta > 0$ be given. If $c^{\bar{k}} = 0$ for some $\bar{k}$ then $\mu^{\bar{k}}$ is an $\varepsilon_k$-optimal solution.

When $c^k \neq 0$ for all $k = 0, 1, 2, \ldots$, by the nonexpansiveness of the projection operation we have

$$\begin{aligned} \|\mu^{k+1} - \mu^*\|^2 &\leq \|\mu^k + s\frac{c^k}{\|c^k\|} - \mu^*\|^2 \\ &= \|\mu^k - \mu^*\|^2 + s^2 - 2s(\mu^* - \mu^k)^T \frac{c^k}{\|c^k\|}. \end{aligned} \tag{14}$$

Let $a_k(\mu^*) = (\mu^* - \mu^k)^T \frac{c^k}{\|c^k\|}$, which is the distance from $\mu^*$ to the supporting hyperplane $H_k = \{\mu \in \mathscr{M} \mid (\mu - \mu^k)^T c^k = 0\}$, that is, $a_k(\mu^*) = \mathrm{dist}(\mu^*, H_k)$.

On the other hand, we suppose that $q(\mu^k) + \varepsilon_k < q^*$, as otherwise $0 \in \partial_{\varepsilon_k} q(\mu^k)$ and $\mu^k$ is an $\varepsilon_k$-optimal solution. Therefore, as $q(\cdot)$ is continuous on $\mathscr{M}$ we can define the level set $L_k^\varepsilon = \{\mu \in \mathscr{M} \mid q(\mu) = q(\mu^k) + \varepsilon_k\}$, which is closed. Hence, the distance $b_k(\mu^*) = \mathrm{dist}(\mu^*, L_k^\varepsilon)$ is well defined.

Since the set $L_k^\varepsilon$ and the point $\mu^*$ lie on the same side of $H_k$ and any segment joining $\mu^*$ with a point of $H_k$ passes through $L_k^\varepsilon$, we have $a_k(\mu^*) \geq b_k(\mu^*)$. Then, from (14) we obtain

$$\|\mu^{k+1} - \mu^*\|^2 \leq \|\mu^k - \mu^*\|^2 + s^2 - 2s b_k(\mu^*). \tag{15}$$

If $b_k(\mu^*) \geq \frac{s}{2}(1+\delta)$ for all $k = 0, 1, 2, \ldots$, then

$$\|\mu^{k+1}-\mu^*\|^2 \le \|\mu^k-\mu^*\|^2-\delta s^2 \le \|\mu^0-\mu^*\|^2-\delta(k+1)s^2, \tag{16}$$

for all $k$.

But $\|\mu^{k+1}-\mu^*\|^2 \ge 0$. Therefore, $\overline{k}$ exists such that

$$b_{\overline{k}}(\mu^*) = \mathrm{dist}(\mu^*, L^{\varepsilon}_{\overline{k}}) < \frac{s}{2}(1+\delta), \tag{17}$$

and, hence, there exists $\overline{\mu} \in L^{\varepsilon}_{\overline{k}}$ with $q(\overline{\mu}) = q(\mu^{\overline{k}}) + \varepsilon_{\overline{k}}$ that verifies $\|\overline{\mu}-\mu^*\| < \frac{s}{2}(1+\delta)$.

□

**Corollary 1.** *If the set $\mathscr{M}^*$ contains a sphere with radius $r > s/2$ and the $\varepsilon$-subgradient method is applied with $\alpha_k = s/\|c^k\|$, then there exists $k^*$ such that $\mu^{k^*}$ is an $\varepsilon_{k^*}$-optimal solution.*

*Proof.* By Proposition 1, for any $\delta > 0$ there exists $\overline{k}$ such that $\overline{\mu} \in L^{\varepsilon}_{\overline{k}}$ where $q(\overline{\mu}) = q(\mu^{\overline{k}}) + \varepsilon_{\overline{k}}$ with $\|\mu^* - \overline{\mu}\| < \frac{s}{2}(1+\delta)$ for any $\mu^* \in \mathscr{M}^*$.

Let $r > s/2$, then we take $\overline{\delta}$ such that $0 < \overline{\delta} < \frac{r-s/2}{s/2}$, for which some $k^*$ must exist such that

$$\mathrm{dist}(\mu^*, L^{\varepsilon}_{k^*}) < \frac{s}{2}(1+\overline{\delta}) < r,$$

for which there exists $\widehat{\mu} \in L^{\varepsilon}_{k^*}$ such that

$$\|\mu^* - \widehat{\mu}\| < \frac{s}{2}(1+\overline{\delta}) < r,$$

then $\widehat{\mu} \in S(\mu^*; r) \subset \mathscr{M}^*$, that is, $q(\widehat{\mu}) = q^*$.

Since $q(\widehat{\mu}) = q(\mu^{k^*}) + \varepsilon_{k^*}$ (by definition of $L^{\varepsilon}_{k^*}$), then $q(\mu^{k^*}) + \varepsilon_{k^*} = q^*$ and $\mu^{k^*}$ is an $\varepsilon_{k^*}$-optimal solution. □

Note that if $\varepsilon_k = 0$ for all $k$, we have Corollary 2 of Theorem 2.1 in [14]. In this work by default $s = 100$.

## 2.2 *Variant of the Constant Step Rule (VCSR)*

Since $c^k$ is an approximate subgradient, there can exist a $k$ such that $c^k \in \partial_{\varepsilon_k} q(\mu^k)$ with $\|c^k\| = 0$, but $\varepsilon_k$ not being sufficiently small. In order to overcome this trouble we have considered the following variant

$$\alpha_k = \frac{s}{\delta + \|c^k\|}, \tag{18}$$

where $s$ and $\delta$ are positive constants. The following proposition shows its kind of convergence (see [7]).

**Proposition 2.** *Let Assumption 1 hold. Let the optimal set $\mathscr{M}^*$ be nonempty. Suppose that a sequence $\{\mu^k\}$ is calculated by the $\varepsilon$-subgradient method given by (10), with the stepsize (18), where $\sum_{k=1}^{\infty} \varepsilon_k < \infty$. Then*

$$q^* - \limsup_{k\to\infty} q_{\varepsilon_k}(\mu^k) < \frac{s}{2}(\delta + C). \tag{19}$$

Note that for very small values of $\delta$ the stepsize (18) is similar to Shor's classical scaling; in contrast, for big values of $\delta$ (with regard to $\sup\{\|c^k\|\}$) the stepsize (18) looks like a constant stepsize. As a result we have chosen by default $\delta = 10^{-12}$ with $s = 100$.

## *2.3 Diminishing Stepsize Rule with Scaling (DSRS)*

It can be seen from the proof of Proposition 1 that at each iteration of (13) the reduction in the distance to the optimal set $\mathscr{M}^*$ is guaranteed only outside a certain neighborhood of that set, with the size of that neighborhood depending on the value of the steplength $s$. Therefore, to obtain standard convergence results it is necessary to require that $s_k$ tends to zero. The reduction of steplengths, however, should not be too rapid. In particular, if the series $\sum_{k=1}^{\infty} s_k$ is convergent then the sequence $\{\mu^k\}$ has a limit, but this limit may lie outside $\mathscr{M}^*$. So for (13), with $\alpha_k = s_k/\|c^k\|$, we have arrived at the classical conditions:

$$s_k > 0, \quad \{s_k\} \to 0 \quad \text{as} \quad k \to \infty, \quad \text{and} \quad \sum_{k=1}^{\infty} s_k = \infty. \tag{20}$$

There are several alternative proofs of convergence of this method for exact subgradients [13, 5]. Below we present our version of the proof of the convergence of the approximate subgradient method, which is based on Theorem 2.2 given by Shor in [14] for exact subgradients (see also [6]).

**Proposition 3.** *Let $\{\varepsilon_k\} \to 0$. Consider the $\varepsilon$-subgradient iteration (13) for $\alpha_k = s_k/\|c^k\|$, where $s_k > 0$ is such that $\lim_{k\to\infty} s_k = 0$ and $\sum_{k=1}^{\infty} s_k = \infty$. Assume that $\mathscr{M}^*$ is closed, bounded, and non-empty. Then either an index $\bar{k}$ exists such that $\mu^{\bar{k}} \in \mathscr{M}^*$ or else*

$$\lim_{k\to\infty} \operatorname{dist}(\mu^k, \mathscr{M}^*) = 0, \qquad \lim_{k\to\infty} q(\mu^k) = q^*, \tag{21}$$

*and* $\lim_{k\to\infty} q_{\varepsilon_k}(\mu^k) = q^*$.

*Proof.* Let $\mu^* \in \mathscr{M}^*$. If there exists a $\bar{k}$ such that $\mu^{\bar{k}} \in \mathscr{M}^*$, the proposition holds. Assume that this $\bar{k}$ does not exist, i.e $\mu^k \notin \mathscr{M}^*$ for all $k$. Then $\mu^k$ can be an $\varepsilon_k$-optimal solution or not. If $\mu^k$ is not an $\varepsilon_k$-optimal solution, like in the proof of Proposition 1 (see (15)), we obtain

$$\|\mu^{k+1} - \mu^*\|^2 \le \|\mu^k - \mu^*\|^2 + s_k^2 - 2 s_k b_k(\mu^*). \tag{22}$$

For a fixed $a > 0$, consider the set $\{\mu \in \mathscr{M} \mid q(\mu) \geq q^* - a\}$ and its boundary $\Gamma_{q^*-a}$. By assumption, the set $\mathscr{M}^*$ is closed and bounded. Thus $\Gamma_{q^*-a}$ is compact, as $q$ is concave over $\mathscr{M}$ convex, and, hence, it is continuous (see Proposition B.9 in [2]).

Since $\{\varepsilon_k\} \to 0$, there exists $N_\varepsilon$, such that for all $k \geq N_\varepsilon$, $a > \varepsilon_k$. Furthermore, $\mathscr{M}^* \cap \Gamma_{q^*-a} = \emptyset$ and there exists a number

$$\rho(a) = \text{dist}(\Gamma_{q^*-a}, \mathscr{M}^*) = \min_{\mu^* \in \mathscr{M}^*, \lambda \in \Gamma_{q^*-a}} \|\lambda - \mu^*\|. \tag{23}$$

Since $\{s_k\} \to 0$, one can find $N_{\rho(a)} \geq N_\varepsilon$ such that for all $k > N_{\rho(a)}$, $s_k < \rho(a)$ and $a > \varepsilon_k$.

If $q(\mu^k) < q^* - a$, then $b_k(\mu^*) > \rho(a)$ and from (22) we have

$$\|\mu^{k+1} - \mu^*\|^2 \leq \|\mu^k - \mu^*\|^2 - \rho(a)s_k, \quad \forall k > N_{\rho(a)}, \tag{24}$$

as by adding the inequalities $s_k^2 < \rho(a)s_k$ and $-2s_k b_k(\mu^*) < -2s_k\rho(a)$ we obtain $s_k^2 - 2s_k b_k(\mu^*) < -\rho(a)s_k$ for all $k > N_{\rho(a)}$.

By adding (24), we have

$$\|\mu^{k+1} - \mu^*\|^2 \leq \|\mu^0 - \mu^*\|^2 - \rho(a)\sum_{i=1}^{k} s_i, \tag{25}$$

and as $\sum_{k=1}^\infty s_k = \infty$, there must exist $N_a > N_{\rho(a)}$ such that for all $\bar{k} \geq N_a$ it holds $q(\mu^{\bar{k}}) \geq q^* - a$.

From here on both cases (when $\mu^k$ is an $\varepsilon_k$-optimal solution and when it is not) are unified. Note that if $\mu^k$ is an $\varepsilon_k$-optimal solution, we have $q(\mu^k) \geq q^* - \varepsilon_k > q^* - a$.

Define $d(a) = \max_{\lambda \in \Gamma_{q^*-a}} \{\min_{\mu^* \in \mathscr{M}^*} \|\lambda - \mu^*\|\}$. If $q(\mu^{\bar{k}}) \geq q^* - a$, then $\min_{\mu^* \in \mathscr{M}^*} \|\mu^{\bar{k}} - \mu^*\| \leq d(a)$.

By the nonexpansiveness of the projection operator for all $k$ it holds

$$\|\mu^{k+1} - \mu^*\| \leq \left\| \left(\mu^k + s_k \frac{c^k}{\|c^k\|}\right) - \mu^* \right\| \leq \|\mu^k - \mu^*\| + s_k. \tag{26}$$

Therefore, for $k = \bar{k}$ we have

$$\min_{\mu^* \in \mathscr{M}^*} \|\mu^{\bar{k}+1} - \mu^*\| \leq \min_{\mu^* \in \mathscr{M}^*} \|\mu^{\bar{k}} - \mu^*\| + s_k \leq d(a) + s_k, \quad \forall \bar{k} \geq N_a. \tag{27}$$

On the other hand, for all $k > N_{\rho(a)}$ with $q(\mu^k) < q^* - a$, by (24), we have

$$\|\mu^{k+1} - \mu^*\| \leq \|\mu^k - \mu^*\|, \tag{28}$$

and hence,

$$\min_{\mu^* \in \mathscr{M}^*} \|\mu^{k+1} - \mu^*\| \leq \min_{\mu^* \in \mathscr{M}^*} \|\mu^k - \mu^*\|. \tag{29}$$

By combining (27) and (29) we obtain

$$\|\mu^{\bar{k}+1} - \mu^*\| \le d(a) + \max_{\bar{k}>N_a}\{s_{\bar{k}}\} \tag{30}$$

for all $\bar{k} > N_a > N_{\rho(a)}$.

Since $d(a) \to 0$ as $a \to 0$, for all $\delta > 0$ there exists $a_\delta$ such that $d(a_\delta) \le \delta/2$.

Next, one can find an index $N_\delta$ such that $q(\mu^k) \ge q^* - a_\delta$ and $s_k \le \delta/2$ for all $k > N_\delta$. Therefore, for all $k > N_\delta$, by (30) we have

$$\min_{\mu^*\in\mathscr{M}^*} \|\mu^k - \mu^*\| \le \delta. \tag{31}$$

This proves that $\lim_{k\to\infty}\left(\min_{\mu^*\in\mathscr{M}^*}\|\mu^k - \mu^*\|\right) = 0$.

By continuity of $q$, we have $\lim_{k\to\infty} q(\mu^k) = q^*$. Moreover, as $\{\varepsilon_k\} \to 0$, by the inequalities (9) we obtain $\lim_{k\to\infty} q_{\varepsilon_k}(\mu^k) = q^*$, which completes the proof. □

An example of such a stepsize is

$$\alpha_k = \frac{s_k}{\|c^k\|}, \text{ with } s_k = s/\widehat{k}, \tag{32}$$

for $\widehat{k} = \lfloor k/m \rfloor + 1$. We use by default $s = 100$ and $m = 5$.

### 2.4 Diminishing Stepsize Rule (DSR)

The convergence of the subgradient method using a diminishing stepsize was shown by Correa and Lemaréchal, see [3]. Next, we consider the special case where $c^k$ is an $\varepsilon_k$-subgradient and $\alpha_k = s_k$ in (13).

The following proposition is proved in [8].

**Proposition 4.** *Let the optimal set $\mathscr{M}^*$ be nonempty. Also, assume that the sequences $\{s_k\}$ and $\{\varepsilon_k\}$ are such that*

$$s_k > 0,\ \sum_{k=0}^{\infty} s_k = \infty,\ \sum_{k=0}^{\infty} s_k^2 < \infty,\ \sum_{k=0}^{\infty} s_k\varepsilon_k < \infty. \tag{33}$$

*Then, the sequence $\{\mu^k\}$, generated by the $\varepsilon$-subgradient method, where $c^k \in \partial_{\varepsilon_k} q(\mu^k)$ (with $\{\|c^k\|\}$ bounded), converges to some optimal solution.*

An example of such a stepsize is

$$\alpha_k = s_k = s/\widehat{k}, \tag{34}$$

for $\widehat{k} = \lfloor k/m \rfloor + 1$. In this work we use by default $s = 100$ and $m = 5$.

An interesting alternative for the ordinary subgradient method is the *dynamic stepsize rule*

$$\alpha_k = \gamma_k \frac{q^* - q(\mu^k)}{\|c^k\|^2}, \tag{35}$$

with $c^k \in \partial q(\mu^k)$ and $0 < \underline{\gamma} \leq \gamma_k \leq \overline{\gamma} < 2$, [13, 14].

Unfortunately, in most practical problems $q^*$ and $q(\mu^k)$ are unknown. Then, the latter can be approximated by $q_{\varepsilon_k}(\mu^k) = l(x^k, \mu^k)$ and $q^*$ replaced with an estimate $q^k_{lev}$. This leads to the stepsize rule

$$\alpha_k = \gamma_k \frac{q^k_{lev} - q_{\varepsilon_k}(\mu^k)}{\|c^k\|^2}, \tag{36}$$

where $c^k \in \partial_{\varepsilon_k} q(\mu^k)$ is bounded for $k = 0, 1, \ldots$.

## *2.5 Dynamic Stepsize with Adjustment Procedure (DSAP)*

An option to estimate $q^*$ is to use the *adjustment procedure* suggested by Nedić and Bertsekas [12], but fitted for the $\varepsilon$-subgradient method

In this procedure $q^k_{lev}$ is the best function value achieved up to the $k$th iteration, in our case $\max_{0 \leq j \leq k} q_{\varepsilon_j}(\mu^j)$, plus a positive amount $\delta_k$, which is adjusted according to algorithm's progress.

The adjustment procedure obtains $q^k_{lev}$ as follows:

$$q^k_{lev} = \max_{0 \leq j \leq k} q_{\varepsilon_j}(\mu^j) + \delta_k, \tag{37}$$

and $\delta_k$ is updated according to

$$\delta_{k+1} = \begin{cases} \rho\delta_k, & \text{if } q_{\varepsilon_{k+1}}(\mu^{k+1}) \geq q^k_{lev}, \\ \\ \max\{\beta\delta_k, \delta\}, & \text{if } q_{\varepsilon_{k+1}}(\mu^{k+1}) < q^k_{lev}, \end{cases} \tag{38}$$

where $\delta_0$, $\delta$, $\beta$, and $\rho$ are fixed positive constants with $\beta < 1$ and $\rho \geq 1$.

The convergence of the approximate subgradient method for this stepsize was analyzed in [8]; see also [6].

# 3 Solution to NCNFP

An algorithm is given below for solving **NCNFP**. This algorithm uses the approximate subgradient method described in Sect. 1.

The value of the dual function $q(\mu^k)$ is estimated by minimizing approximately $l(x,\mu^k)$ over $x \in \mathscr{F}$ (the set defined by the network constraints) so that the optimality tolerance, $\tau_x^k$, becomes more rigorous as $k$ increases, i.e. the minimization will be *asymptotically exact* [1]. In other words, we set $q_{\varepsilon_k}(\mu^k) = l(x^k,\mu^k)$, where $x^k$ minimizes approximately the nonlinear network subproblem $\mathbf{NNS_k}$

$$\underset{x \in \mathscr{F}}{\text{minimize}}\ l(x,\mu^k) \tag{39}$$

in the sense that this minimization stops when we obtain an $x^k$ that verifies the KKT conditions with $\tau_x^k$ accuracy, which implies that the norm of the reduced gradient holds

$$\|Z^T \nabla_x l(x^k,\mu^k)\| \leq \tau_x^k, \tag{40}$$

where $\lim_{k\to\infty} \tau_x^k = 0$ and $Z$ represents the reduction matrix whose columns form a base of the null subspace generated by the rows of the matrix of active network constraints of this subproblem (including the active capacity constraints on the flows of each arc), see [11]. Let $\bar{x}^k$ be the minimizer of this subproblem approximated by $x^k$. Then, it can be proved (see [8]) that there exists a positive $w$, such that $l(x^k,\mu^k) \leq l(\bar{x}^k,\mu^k) + w\tau_x^k$ for $k = 1,2,\ldots$. If we set $\varepsilon_k = \omega\tau_x^k$, this inequality becomes (8). Moreover, as

$$\tau_x^{k+1} = \sigma\tau_x^k, \qquad \text{for a fixed} \quad \sigma \in (0,1), \tag{41}$$

then $\sum_{k=1}^{\infty} \varepsilon_k < \infty$, and so $\lim_{k\to\infty} \varepsilon_k = 0$. Consequently, to solve this problem we can use the approximate subgradient methods with the stepsizes described in Sect. 2. We denote $q_{\varepsilon_k}(\mu^k) = l(x^k,\mu^k)$, which satisfies the inequality (9). In this work, $\sigma = 10^{-1}$ by default. Note that in this case, $\varepsilon_k = \tau_x^k\omega = 10^{-(k-1)}\tau_x^1\omega$.

**Algorithm 3.1 (Approximate subgradient method for NCNFP)**

**Step 0** *Initialize.* Set $k = 1$, $N_{max}$, $\tau_x^1$, $\varepsilon_q$, $\varepsilon_\mu$ and $\tau_\mu$. Set $\mu^1 = 0$.
**Step 1** *Compute* the dual function estimate, $q_{\varepsilon_k}(\mu^k)$, by solving $\mathbf{NNS_k}$ with accuracy $\tau_x^k$, then $x^k \in \mathscr{F}$ is an approximate solution, $q_{\varepsilon_k}(\mu^k) = l(x^k,\mu^k)$, and $c^k = c(x^k)$ is an $\varepsilon_k$-subgradient of $q$ in $\mu^k$.
**Step 2** *Check the stopping rules* for $\mu^k$.

$T_1$: Stop if $\max\limits_{i=1,\ldots,r} \left\{(c_i^k)^+\right\} < \tau_\mu$ , where $(c_i^k)^+ = \max\{0, c_i(x^k)\}$.

$T_2$: Stop if $\dfrac{|q^k - (q^{k-1} + q^{k-2} + q^{k-3})/3|}{1 + |q^k|} < \varepsilon_q$ , where $q^l = q_{\varepsilon_l}(\mu^l)$.

$T_3$: Stop if $\dfrac{1}{5}\sum\limits_{i=0}^{4} \|\mu^{k-i} - \mu^{k-i-1}\|_\infty < \varepsilon_\mu$.

$T_4$: Stop if $k$ reaches a prefixed value $N_{max}$.

If $\mu^k$ fulfils one of these tests, then it is deemed approximately optimal, and $(x^k,\mu^k)$ is an approximate primal-dual solution.

**Step 3** *Update* the estimate $\mu^k$ by means of the iteration

$$\mu_i^{k+1} = \begin{cases} \mu_i^k + \alpha_k c_i^k, \text{ if } \mu_i^k + \alpha_k c_i^k > 0 \\ 0, \qquad\qquad \text{otherwise} \end{cases} \tag{42}$$

where $\alpha_k$ is computed using some stepsize rule. Go to Step 1.

In Step 0, for the stopping rules, $\tau_\mu = 10^{-5}$, $\varepsilon_q = 10^{-7}$, $\varepsilon_\mu = 10^{-3}$ and $N_{max} = 200$ have been taken. In addition, $\tau_x^1 = 10^{-2}$ by default.

Step 1 is carried out by the code PFNL (described in [9]), which is based on the specific procedures for nonlinear network flows [15] and the active set procedure [11], using a spanning tree as the basis matrix of the network constraints.

In Step 2, alternative heuristic tests have been used for practical purposes. $T_1$ checks the feasibility of $x^k$, as if the violation of the side constraints has been sufficiently reduced, then $(x^k, \mu^k)$ is an acceptable primal-dual solution for **NCNFP**. $T_2$ and $T_3$ mean that $\mu$ does not improve for the last iterations. $T_4$ is used to stop the algorithm when this is not able to find a good enough solution.

To obtain $\alpha_k$ in Step 3, we have used the iteration (10) (see Sect. 1) with the five stepsize rules considered in Sect. 2: CSR, VCSR, DSRS, DSR, and DSAP. In the implementation of DSAP we use $\rho = 2$, $\beta = 1/\rho$, $\delta_0 = 0.5\|(c^1)^+\|$, and $\delta = 10^{-7}|l(x^0, \mu^1)|$, where $x^0$ is the initial feasible point for Step 1 and $k = 1$.

The values given above have been heuristically chosen. The implementation in Fortran-77 of the previous algorithm, termed PFNRN05, was designed to solve large-scale nonlinear network flow problems with nonlinear side constraints.

## 4 Numerical Tests

In order to obtain a computational comparison of the performance of the stepsizes CSR, VCSR, DSRS, DSR, and DSAP, some computational tests are carried out, which consist in solving nonlinear network flow problems with nonlinear side constraints using PFNRN05 code with the alternative stepsizes, where the objective functions are strictly convex and the side-constraint functions are convex. Therefore, these problems have a unique primal solution $x^*$ and the duality gap is zero. The numerical tests have been carried out on a Sun Enterprise 250 under UNIX.

The problems used in these tests were created by means of the DIMACS-random-network generators Rmfgen and Gridgen (see [4]). These generators provide linear flow problems in networks without side constraints. The side constraints are defined by convex quadratic functions and were generated through the *Dirnl* random generator described in [9, 10].

These test problems have up to 4008 variables, 1200 nodes, and 1253 side constraints, see [7]. The objective functions are nonlinear and strictly convex, and are either Namur functions (**n1**) or polynomial functions (**e2**). The polynomial functions give rise to problems with a moderate number of superbasic variables (degrees of freedom) at the optimizer, whereas the Namur functions [15] generate a high number of superbasic variables. More details in [7].

In Table 1 we compare the quality of the solution by means of the value of maximum violation of the side constraints at the optimal solution, $c^* = \|[c(x^*)]^+\|_\infty$, and the efficiency by the CPU times (in seconds) used to compute the solution. Note that $c^*$ offers information about the feasibility of this solution and, hence, about its duality gap.

**Table 1** Comparison of the quality/efficiency for the stepsizes

| Prob. | CSR | VCSR | DSRS | DSR | DSAP |
|---|---|---|---|---|---|
| c15e2 | $10^{-5}/4.5$ | $10^{-5}/4.4$ | $10^{-4}/8.5$ | $10^{-4}/2.9$ | $10^{-8}/2.2$ |
| c17e2 | $10^{-5}/6.0$ | $10^{-5}/7.2$ | $10^{-4}/13.7$ | $10^{-4}/6.3$ | $10^{-6}/3.1$ |
| c18e2 | $10^{-5}/21.3$ | $10^{-5}/30.7$ | $10^{-4}/67.0$ | $10^{-2}/24.3$ | $10^{-5}/38.0$ |
| c13n1 | $10^{-6}/64.3$ | $10^{-6}/54.5$ | $10^{-6}/64.3$ | $10^{-5}/84.6$ | $10^{-6}/44.2$ |
| c15n1 | $10^{-6}/82.3$ | $10^{-6}/61.4$ | $10^{-6}/83.2$ | $10^{-4}/426.9$ | $10^{-8}/99.8$ |
| c17n1 | $10^{-5}/78.9$ | $10^{-5}/75.2$ | $10^{-5}/79.1$ | $10^{-4}/361.1$ | $10^{-6}/92.3$ |
| c22e2 | $10^{-5}/3.8$ | $10^{-5}/3.9$ | $10^{-4}/8.5$ | $10^{-4}/2.5$ | $10^{-6}/2.0$ |
| c23e2 | $10^{-5}/6.3$ | $10^{-5}/6.2$ | $10^{-4}/6.0$ | $10^{-3}/5.0$ | $10^{-7}/5.9$ |
| c24e2 | $10^{-5}/15.9$ | $10^{-5}/15.9$ | $10^{-4}/94.8$ | $10^{-2}/40.4$ | $10^{-9}/5.2$ |
| c34e2 | $10^{-6}/4.8$ | $10^{-6}/5.3$ | $10^{-6}/4.9$ | – | $10^{-8}/5.1$ |
| c35e2 | $10^{-6}/2.0$ | $10^{-6}/2.3$ | $10^{-6}/2.0$ | $10^{-7}/2.9$ | $10^{-9}/2.4$ |
| c38e2 | $10^{-6}/8.2$ | $10^{-5}/7.4$ | $10^{-6}/8.4$ | $10^{-5}/12.4$ | $10^{-8}/8.0$ |
| c42e2 | $10^{-7}/19.0$ | $10^{-7}/14.1$ | $10^{-7}/18.9$ | $10^{-7}/14.1$ | $10^{-7}/14.5$ |
| c47e2 | $10^{-5}/404.9$ | $10^{-5}/401.0$ | $10^{-5}/448.0$ | – | $10^{-7}/657.1$ |
| c48e2 | $10^{-5}/114.4$ | $10^{-5}/132.2$ | $10^{-5}/116.7$ | – | $10^{-7}/257.6$ |

Table 1 points out that the quality of the solution computed by PFNRN05 for the stepsize DSR is lower than that of DSRS, whereas that of this stepsize is slightly lower than that of CSR and VCSR. Also, the quality of the solution obtained with the stepsize DSAP is clearly higher than that computed for the rest of stepsizes.

Regarding the efficiency of PFNRN05 for these stepsizes, we observe that the efficiency when we use DSAP is similar to that obtained for CSR, VCSR, and DSRS, whereas our code for DSR is less efficient and robust than for the other stepsizes.

## 5 Conclusions

In this work some basic convergence results of subgradient methods for the stepsizes CSR and DSRS have been extended to approximate subgradient methods. Moreover, in the numerical tests carried out over convex nonlinear problems of nonlinearly constrained networks we have observed that the quality of the solution obtained by PFNRN05 for the dynamic stepsize rule DSAP is higher than that obtained for the other stepsizes, while the efficiency is similar.

The results of the numerical tests encourage to carry out further experimentation with other kind of problems and to compare the efficiency with that of well-known codes.

## References


1. D. P. Bertsekas: *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press, New York (1982)
2. D. P. Bertsekas: *Nonlinear Programming*. 2nd ed. Athena Scientific, Belmont, Massachusetts (1999)
3. R. Correa, C. Lemarechal: Convergence of some algorithms for convex minimization. *Mathematical Programming* **62** pp. 261–275 (1993)
4. DIMACS. The first DIMACS international algorithm implementation challenge: The benchmark experiments. Technical Report, DIMACS, New Brunswick, NJ, USA (1991)
5. Yu. M. Ermoliev: Methods for solving nonlinear extremal problems. *Cybernetics* **2** pp. 1–17 (1966)
6. K. Kiwiel: Convergence of approximate and incremental subgradient methods for convex optimization. *SIAM Journal on Optimization* **14**(3) pp. 807–840 (2004)
7. E. Mijangos: A variant of the constant step rule for approximate subgradient methods over nonlinear networks. M. Gavrilova et al. (eds.): ICCSA 2006, *Lecture Notes in Computer Science* **3982** pp. 757-766, Springer-Verlag, Berlin Heidelberg (2006)
8. E. Mijangos: Approximate subgradient methods for nonlinearly constrained network flow problems. *Journal of Optimization Theory and Applications* **128**(1) pp. 167–190 (2006)
9. E. Mijangos, N. Nabona: *The application of the multipliers method in nonlinear network flows with side constraints*. Technical Report 96/10, Dept. of Statistics and Operations Research, Universitat Politècnica de Catalunya, Barcelona, Spain (1996) (downloadable from website http://www.ehu.es/∼mepmifee).
10. E. Mijangos, N. Nabona: On the first-order estimation of multipliers from Kuhn-Tucker systems. *Computers and Operations Research* **28** pp. 243–270 (2001)
11. B. A. Murtagh, M. A. Saunders: Large-scale linearly constrained optimization. *Mathematical Programming* **14** pp. 41–72 (1978)
12. A. Nedić, D. P. Bertsekas: Incremental subgradient methods for nondifferentiable optimization. *SIAM Journal on Optimization* **12** pp. 109–138 (2001)
13. B. T. Poljak: *Introduction to Optimization*, Optimization Software Inc., New York (1987)
14. N. Z. Shor: *Minimization Methods for Nondifferentiable Functions*. Springer-Verlag, Berlin (1985)
15. Ph. L. Toint, D. Tuyttens: On large scale nonlinear network optimization. *Mathematical Programming* **48** pp. 125–159 (1990)