March 8, 1989 LIDS-P-1859

A Decomposition Algorithm for Convex Differentiable Minimization*

by

Paul Tseng†

Abstract

In this paper we propose a decomposition algorithm for convex differentiable minimization. This algorithm at each iteration solves a variational inequality problem obtained by approximating the gradient of the cost function by a strongly monotone function. A line search is then performed in the direction of the solution to this variational inequality (with respect to the original cost). If the constraint set is a Cartesian product of m sets, the variational inequality decomposes into m coupled variational inequalities which can be solved in either a Jacobi manner or a Gauss-Seidel manner. This algorithm also applies to the minimization of strongly convex (possibly nondifferentiable) costs subject to linear constraints. As special cases, we obtain the GP-SOR algorithm of Mangasarian and De Leone, a diagonalization algorithm of Feijoo and Meyer, the coordinate descent method, and the dual gradient method. This algorithm is also closely related to a splitting algorithm of Gabay and a gradient projection algorithm of Goldstein and Levitin-Poljak, and has interesting applications to separable convex programming and to solving traffic assignment problems.

KEY WORDS: convex programming, decomposition, linear complementarity, variational inequality.

---

* This research is partially supported by the U.S. Army Research Office, contract DAAL03-86-K-0171 (Center for Intelligent Control Systems), and by the National Science Foundation under grant NSF-ECS-8519058.

† The author is with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139

1. Introduction

In convex differentiable minimization, one frequently encounters problems whose solution simplifies considerably if the cost functions were separable. Examples of this include problems whose constraint sets have product forms (such as the traffic assignment problem) or are polyhedral [Roc83]. In this case, it is desirable to approximate the original cost function by a sequence of separable cost functions. The classical gradient descent method is one example of a method that follows this approach (approximating the original cost by a sequence of linear costs), but it suffers from slow convergence. Another example is the coordinate descent method, but the convergence of this method requires the cost to be in some sense strictly convex and in general applies to only the Gauss-Seidel version. Recently, Feijoo and Meyer [FeM88] (also see [LiP87] for the quadratic case) proposed a Jacobi version of the coordinate descent method that circumvents the difficulties with convergence by introducing a line search at each iteration. Also recently, Mangasarian and De Leone [MaD88] proposed a matrix splitting method for solving symmetric linear complementarity problems that also introduces a line search step at each iteration. In this paper we show that these two methods may be viewed naturally as special cases of a Jacobi-type feasible descent method. This Jacobi method, at each iteration, uses as the descent direction the solution to a variational inequality problem obtained by replacing the gradient of the cost function by a strongly monotone continuous function. A line search (possibly inexact) is then performed along this direction. A major advantage of this method is that each strongly monotone function can be chosen arbitrarily; hence it can be chosen either to match the structure of the constraint set or to match the structure of the cost function. Furthermore, when the constraint set is a Cartesian product, it can be implemented in a Gauss-Seidel manner (thus accelerating the convergence rate). A special case of this Gauss-Seidel method is the classical coordinate descent method [D'Es59], [Lue84], [Pow73], [SaS73], [Zan69]. It can also be implemented as a dual method for minimizing strongly convex (possibly nondifferentiable) functions subject to linear constraints. A special case of this dual method is the dual gradient method [Pan86, §6]. This algorithm is also closely related to a splitting algorithm of Gabay [Gab83] and a gradient projection algorithm of Goldstein and Levitin-Poljak [Gol64], [LeP66] – the main difference being that an additional line search is used at every iteration.

This paper proceeds as follows: In §2, we describe the Jacobi-type feasible descent method and establish its convergence. In §3 we give a Gauss-Seidel version of this method for problems whose constraint set is a Cartesian product. In §4 we give a dual version of this method for minimizing strongly convex functions subject to linear constraints. In §5 we study the relationship between the new method and those known and propose applications to separable cost problems and the solution of traffic assignment problems. Finally, in §6 we discuss possible extensions.

In our notation, all vectors are column vectors and superscript T denotes transpose. We denote by $\langle\cdot,\cdot\rangle$ the usual Euclidean inner product and by $\|\cdot\|$ its induced norm. For any set S in $\Re^n$ ($n \geq 1$), we denote by cl(S) the closure of S and ri(S) the relative interior of S. We use $\Re^n_+$ to denote the nonnegative orthant in $\Re^n$. For any closed convex set $S \subseteq \Re^n$, we denote by $[\cdot]_S^+$ the orthogonal projection onto S. For any convex function $f:\Re^n \rightarrow (-\infty,\infty]$, we denote by dom(f) its effective domain, by $\partial f(x)$ its subdifferential at x and by $f'(x;d)$ its directional derivative at x in the direction d. We also denote by f^* the <u>conjugate</u> function of f [Roc70], i.e.

$$f^*(y) \;=\; \sup_x \; \{ \langle y, x\rangle - f(x) \}, \qquad \forall\; y \in \Re^n.$$

Finally, for any closed set $S \subseteq \Re^n$ and any function $F:S \rightarrow \Re^n$, we say that F is <u>strictly monotone</u> if

$$\langle F(y) - F(x), y - x\rangle \;>\; 0, \quad \forall\; x \in S, \forall\; y \in S.$$

Similarly, we say that F is <u>strongly monotone</u> with modulus (α,σ), for some $\alpha > 0$ and $\sigma > 1$, if

$$\langle F(y) - F(x), y - x\rangle \;\geq\; \alpha\|y-x\|^\sigma, \quad \forall\; x \in S, \forall\; y \in S. \tag{1.1}$$

2. A Jacobi-Type Feasible Descent Algorithm

Consider the following convex program

$$\begin{array}{ll} \text{Minimize} & f(x) \\ \text{subject to} & x \in X, \end{array} \tag{2.1}$$

where X is a nonempty closed convex set in $\Re^n$ and $f:X\to\Re$ is a convex function. We make the following standing assumptions:

Assumption A:

(a) f is convex and continuously differentiable on X.

(b) f is bounded from below on X.

Note that we do not assume that (2.1) has an optimal solution. [We remark that our results also extend to the more general case where f(x) is allowed to tend to $+\infty$ as x approaches the boundary of X. This is because the method that we propose is a feasible descent method so that we can in effect replace X by its intersection with some level set of f (which is a closed convex set). The function f is continuously differentiable on this intersection.]

Consider the following feasible descent method for solving (2.1), whereby at each iteration we solve a strongly monotone variational inequality problem to generate the descent direction:

NPPD Algorithm:

Iter. 0 Choose any $\alpha > 0$, $\sigma > 1$, and $x^1 \in X$. Also choose any continuous function $W:X\times X\to\Re^n$ such that $W(\cdot,x)$ is strongly monotone with modulus (α,σ) for each $x\in X$.

Iter. r Compute y^r to be the unique $x\in X$ satisfying the variational inequality:

$$\langle W(x,x^r) - W(x^r,x^r) + \nabla f(x^r), y - x\rangle \geq 0, \quad \forall\, y\in X, \tag{2.2}$$

and perform a line search along the direction $y^r - x^r$ from x^r:

$$x^{r+1} \leftarrow x^r + \theta^r(y^r - x^r), \tag{2.3}$$

where $\theta^r = \operatorname{argmin}\{\ f(x^r + \theta(y^r - x^r)) \mid x^r + \theta(y^r - x^r) \in X\ \}$.

We have called the above algorithm the NPPD (for Nonlinear Proximal Point Descent) algorithm because, in the absence of the line search step, it looks like a nonlinear version of the <u>proximal point</u> algorithm [Luq86], [Mar70], [Roc76b] (also see Example 2 in §5).

The choice of the function W is quite crucial in determining the efficiency of this algorithm. For example, if $W(\cdot,x)$ is chosen as the gradient of a strongly convex differentiable function defined on X, then (2.2) can be solved as a convex program. [In general, we can use either the gradient projection method [BeT89, §3.5], or the extragradient method [Kor76], or a certain splitting algorithm [LiM79] to solve (2.2).] Moreover, if X has a decomposable structure, then we can choose W to match this structure and thus simplifying the computation. As an example, suppose that $X = X_1\times\ldots\times X_m$ for some closed convex sets $X_1\subseteq\Re^{n_1}, \ldots, X_m\subseteq\Re^{n_m}$. By choosing

$$W(x_1,\ldots,x_m,y) = (W_1(x_1,y),\ W_2(x_2,y),\ \ldots,\ W_m(x_m,y)), \qquad (2.4)$$

where $x_i\in X_i$ and each $W_i:X_i\times X\to\Re^{n_i}$ is a continuous function that is strongly monotone in x_i, we decompose the variational inequality (2.2) into m independent variational inequalities which can be solved in parallel. One possible choice for W_i is

$$W_i(x_i,y) = \nabla_i f(y_1,\ldots,y_{i-1},x_i,y_{i+1},\ldots,y_m) + \rho(x_i-y_i),$$

where $\nabla_i f$ denotes the partial derivative of f with respect to x_i and ρ is any positive scalar (if f is strongly convex, then $\rho = 0$ is also acceptable). In fact, the Feijoo-Meyer method can be seen to correspond to the above choice of W_i with $\rho = 0$. [However, convergence of the Feijoo-Meyer method does not require f to be strongly convex.] Also we remark that the stepsize θ^r does not have to be computed exactly. It suffices to use any $\theta\in(0,\bar{\theta}^r]$ satisfying

$$\beta\langle\nabla f(x^r),y^r-x^r\rangle \ \le\ \langle\nabla f(x^r+\theta(y^r-x^r)),y^r-x^r\rangle \ \le\ 0,$$

if such θ exits, and to use $\theta = \bar{\theta}^r$ otherwise, where β is a fixed scalar in (0,1) and $\bar{\theta}^r$ is the largest θ for which $x^r + \theta(y^r-x^r)\in X$. For example, we can use a variant of the Armijo rule [Ber82] to compute such a stepsize.

We show below that the sequence of iterates generated by the NPPD algorithm is in some sense convergent:

Proposition 1 Let $\{x^r\}$ be a sequence of iterates generated by the NPPD algorithm. Then every limit point of $\{x^r\}$ is an optimal solution of (2.1).

Proof: Fix any integer $r \ge 0$ and let y^r be the solution of (2.2). Then

$$\langle W(y^r,x^r)-W(x^r,x^r)+\nabla f(x^r),y-y^r\rangle \ \ge\ 0, \quad \forall\ y\in X. \tag{2.5}$$

Hence (cf. (1.1))

$$\begin{aligned} 0 \ &\le\ \langle W(y^r,x^r)-W(x^r,x^r)+\nabla f(x^r),x^r-y^r\rangle \\ &\le\ -\alpha\|y^r-x^r\|^\sigma + \langle\nabla f(x^r),x^r-y^r\rangle, \end{aligned} \tag{2.6}$$

so that y^r-x^r is a descent direction at x^r. Suppose that x^∞ is a limit point of $\{x^r\}$ and let $\{x^r\}_{r\in R}$ be a subsequence converging to x^∞. Since $W(\cdot,x^r)$ is strongly monotone, this implies that $\{y^r\}_{r\in R}$ is bounded. [For any $y\in\Re^n$ such that $\alpha\inf_{r\in R}\{\|y-x^r\|^{\sigma-1}\} > \sup_{r\in R}\{\|\nabla f(x^r)\|\}$, we have $\langle W(y,x^r)-W(x^r,x^r)+\nabla f(x^r),x^r-y\rangle < 0$ for all $r\in R$.] By further passing into a subsequence if necessary, we can assume that $\{y^r\}_{r\in R}$ converges to some limit point y^∞. We claim that $x^\infty = y^\infty$. To see this, note that (2.6) and the continuity of ∇f on X implies that

$$\langle\nabla f(x^\infty),\ y^\infty-x^\infty\rangle \ \le\ -\alpha\|y^\infty-x^\infty\|^\sigma.$$

Hence if $y^\infty \ne x^\infty$, then $\langle\nabla f(x^\infty),y^\infty-x^\infty\rangle < 0$ and there exists an $\varepsilon\in(0,1]$ such that

$$\langle\nabla f(x^\infty + \varepsilon(y^\infty-x^\infty)),\ y^\infty-x^\infty\rangle \ \le\ .5\langle\nabla f(x^\infty),y^\infty-x^\infty\rangle.$$

Since $\{x^r\}_{r\in R} \to x^\infty$ and $\{y^r\}_{r\in R} \to y^\infty$, this implies

$$\langle\nabla f(x^r + \varepsilon(y^r-x^r)),\ y^r-x^r\rangle \ \le\ .4\langle\nabla f(x^\infty),\ y^\infty-x^\infty\rangle,$$

for all $r\in R$ sufficiently large. This in turn implies that

$$f(x^r + \varepsilon(y^r-x^r)) - f(x^r) \ \le\ .4\varepsilon\langle\nabla f(x^\infty),\ y^\infty-x^\infty\rangle,$$

and (cf. (2.3)) $\theta^r \geq \varepsilon$. Hence, the quantity $f(x^{r+1}) - f(x^r)$ is bounded from above by a negative scalar constant, for all $r \in R$ sufficiently large. Since $f(x^r)$ is monotonically decreasing with r, it must be that $f(x^r) \to -\infty$, a contradiction of Assumption A (b).

Since $\{x^r\}_{r\in R} \to x^\infty$, $\{y^r\}_{r\in R} \to x^\infty$, by passing into the limit in (2.5) and using the continuity of W, we obtain that

$$\langle \nabla f(x^\infty), x - x^\infty \rangle \geq 0, \quad \forall\, x \in X.$$

Since f is convex, this implies that x^∞ is an optimal solution of (2.1). Q.E.D.

As a corollary of Proposition 1, we obtain that if f has bounded level sets on X, then $\{x^r\}$ is bounded and each one of its limit points solves (2.1). If X is itself bounded, then the strong monotonicity assumption on W can be weakened somewhat:

Proposition 1′ Assume that X is bounded and let $\{x^r\}$ be a sequence of iterates generated by the NPPD algorithm. Then, even if $W(\cdot,x)$ is only strictly monotone (instead of strongly monotone) for each $x \in X$, every limit point of $\{x^r\}$ is an optimal solution of (2.1).

Proof: The boundedness of X implies that both $\{x^r\}$ and the sequence $\{y^r\}$ given by (2.2) are bounded. Hence for each limit point x^∞ of $\{x^r\}$ we can find a subsequence $\{x^r\}_{r\in R}$ and a $y^\infty \in X$ such that $\{x^r\}_{r\in R} \to x^\infty$ and $\{y^r\}_{r\in R} \to y^\infty$. Letting $y = x^\infty$ in (2.2) and passing into the limit as $r \to \infty$, $r \in R$ (also using the continuity of W), we obtain

$$\langle \nabla f(x^\infty), y^\infty - x^\infty \rangle \leq -\langle W(y^\infty, x^\infty) - W(x^\infty, x^\infty), y^\infty - x^\infty \rangle.$$

Since $W(\cdot, x^\infty)$ is strictly monotone, if $y^\infty \neq x^\infty$, the above inequality would imply $\langle \nabla f(x^\infty), y^\infty - x^\infty \rangle < 0$ and hence (by an argument analogous to that in the proof of Proposition 1) $f(x^r) \to -\infty$, a contradiction of Assumption A (b). Therefore $y^\infty = x^\infty$.

Since $\{x^r\}_{r\in R} \to x^\infty$, $\{y^r\}_{r\in R} \to x^\infty$, by passing into the limit in (2.5) and using the continuity of W, we obtain that

$$\langle \nabla f(x^\infty), x - x^\infty \rangle \geq 0, \quad \forall\, x \in X,$$

so that x^∞ is an optimal solution of (2.1).   Q.E.D.

In general, the convergence of $\{x^r\}$ remains an open problem. In the special case where f is Lipschitz continuous on X we can show that $\{x^r\}$ is in some sense approaching the optimal solution set (recall that x is an optimal solution of (2.1) if and only if $[\, x - \nabla f(x)\,]_X^+ = x$):

Proposition 2 Assume that ∇f is Lipschitz continuous on X and let $\{x^r\}$ be a sequence of iterates generated by the NPPD algorithm. If furthermore W is uniformly continuous on X×X, then $[x^r - \nabla f(x^r)]_X^+ - x^r \to 0$.

Proof: Since ∇f is Lipschitz continuous on X, there exists $\lambda > 0$ and $\eta > 0$ such that

$$\|\nabla f(y) - \nabla f(x)\| \leq \lambda \|y - x\|^\eta, \quad \forall\, x \in X,\ \forall\, y \in X.$$

Hence the directional derivative $\langle \nabla f(x), y^r - x^r \rangle$ can increase by at most $\lambda \cdot \theta^\eta \cdot \|y^r - x^r\|^{1+\eta}$ when x is moved from x^r along the direction $y^r - x^r$ by an amount $\theta > 0$. This implies that (cf. (2.3))

$$\theta^r \geq \min\{\, 1, (\langle \nabla f(x^r), x^r - y^r \rangle / (\lambda \cdot \|y^r - x^r\|^{1+\eta}))^{1/\eta} \}. \tag{2.7}$$

Also we have, for all $\theta' \in [0, \theta^r]$,

$$\begin{aligned} f(x^r + \theta'(y^r - x^r)) - f(x^r) &= \int_0^{\theta'} \langle \nabla f(x^r + \theta(y^r - x^r)), y^r - x^r \rangle \, d\theta \\ &\leq \int_0^{\theta'} \langle \nabla f(x^r), y^r - x^r \rangle + \lambda \theta^\eta \|y^r - x^r\|^{1+\eta} \, d\theta \\ &= \langle \nabla f(x^r), y^r - x^r \rangle \theta' + \lambda (\theta')^{\eta+1} (\eta+1)^{-1} \|y^r - x^r\|^{1+\eta}. \end{aligned} \tag{2.8}$$

If $\langle \nabla f(x^r), x^r - y^r \rangle \geq \lambda \cdot \|y^r - x^r\|^{1+\eta}$, then (cf. (2.7)) $\theta^r \geq 1$, so that (2.8) with $\theta' = 1$ implies

$$f(x^{r+1}) - f(x^r) \leq -\lambda \cdot \eta\, (\eta+1)^{-1} \cdot \|y^r - x^r\|^{1+\eta}.$$

Otherwise, $\theta^r \geq (\langle \nabla f(x^r), x^r - y^r \rangle / (\lambda \cdot \|y^r - x^r\|^{1+\eta}))^{1/\eta}$ so that (2.8) with $\theta' = (\langle \nabla f(x^r), x^r - y^r \rangle / (\lambda \cdot \|y^r - x^r\|^{1+\eta}))^{1/\eta}$ implies

$$f(x^{r+1}) - f(x^r) \leq -[\langle \nabla f(x^r), x^r - y^r \rangle / \|y^r - x^r\|]^{1+1/\eta} \cdot \eta(\eta+1)^{-1} \cdot \lambda^{-1/\eta},$$

which, together with (2.6), implies

$$f(x^{r+1}) - f(x^r) \leq -[\alpha \|y^r - x^r\|^{\sigma-1}]^{1+1/\eta} \cdot \eta(\eta+1)^{-1} \cdot \lambda^{-1/\eta}.$$

Hence, in either case, $y^r - x^r \to 0$. Now

$$\begin{aligned} \|[x^r - \nabla f(x^r)]_X^+ - x^r\| &= \|[x^r - \nabla f(x^r)]_X^+ - [y^r - W(y^r,x^r) + W(x^r,x^r) - \nabla f(x^r)]_X^+ \\ &\quad + [y^r - W(y^r,x^r) + W(x^r,x^r) - \nabla f(x^r)]_X^+ - y^r + y^r - x^r\| \\ &\leq \|[x^r - \nabla f(x^r)]_X^+ - [y^r - W(y^r,x^r) + W(x^r,x^r) - \nabla f(x^r)]_X^+\| \\ &\quad + \|[y^r - W(y^r,x^r) + W(x^r,x^r) - \nabla f(x^r)]_X^+ - y^r\| + \|y^r - x^r\| \\ &\leq \|x^r - y^r + W(y^r,x^r) - W(x^r,x^r)\| + 0 + \|y^r - x^r\|, \end{aligned}$$

where the first inequality follows from the triangle inequality and the second inequality follows from (2.2) and the fact that the projection mapping $[\cdot]_X^+$ is nonexpansive (with respect to the L_2-norm). Since $y^r - x^r \to 0$ and W is uniformly continuous, this proves our claim. Q.E.D.

Note that the conclusion of Proposition 2 holds even if (2.1) does not have an optimal solution. Proposition 2 is useful in certain dual applications for which f does not have bounded level sets (see §4 and Examples 3 and 7 in §5).

3. A Gauss-Seidel Algorithm

In this section we consider the convex program (2.1) again, but in addition to Assumption A we make the following assumption:

Assumption B: $X = X_1 \times \ldots \times X_m$, where each X_i is some closed convex set of $\Re^{n_i}$ $(n_1 + \ldots + n_m = n)$.

As we noted in §2, if we choose W in the NPPD algorithm to have the separable form (2.4), then at each iteration we solve simultaneously m variational inequalities defined on $X_1, X_2, \ldots, X_m$ respectively. It is natural then to consider a variant of the NPPD algorithm whereby we solve these m variational inequalities sequentially and use the solution to the previous variational inequalities to construct the current variational inequality. Intuitively, such a Gauss-Seidel variant should converge faster. Below we describe this method and analyze its convergence:

GS-NPPD Algorithm:

Iter. 0 Choose any $\alpha > 0$, $\sigma > 1$, $\bar{\theta} > 0$, and $x_i^1 \in X_i$ $(i = 1,\ldots,m)$. For each $i \in \{1,2,\ldots,m\}$, choose a continuous function $W_i: X_i \times X \to \Re^{n_i}$ such that $W_i(\cdot,x)$ is strongly monotone with modulus (α,σ) for all $x \in X$.

Iter. r Choose $i \in \{1,2,\ldots,m\}$. Compute y_i^r to be the unique $x_i \in X_i$ satisfying the variational inequality

$$\langle W_i(x_i,x^r) - W_i(x_i^r,x^r) + \nabla_i f(x^r), y_i - x_i \rangle \geq 0, \quad \forall\, y_i \in X_i, \tag{3.1}$$

where $x^r = (x_1^r,\ldots,x_m^r)$ and $\nabla_i f$ denotes the partial derivative of f with respect to x_i. Let $d^r = y_i^r - x_i^r$ and compute

$$x_i^{r+1} \leftarrow x_i^r + \theta^r d^r,$$
$$x_j^{r+1} \leftarrow x_j^r, \qquad \forall\, j \neq i,$$

where $\theta^r = \mathrm{argmin}\{\, f(x_1^r,\ldots,x_{i-1}^r,x_i^r+\theta d^r,x_{i+1}^r,\ldots,x_m^r) \mid x_i^r+\theta d^r \in X_i,\ \theta \leq \bar{\theta}\ \}$.

[Note that if $m = 1$ and $\bar{\theta} = \infty$, then the GS-NPPD algorithm reduces to the NPPD algorithm. Also, like the NPPD algorithm, the line search can be inexact.] Let $\sigma(r)$ denote the index i chosen at the rth iteration. To ensure convergence of the GS-NPPD algorithm, we impose the following rule on the sequence $\{\sigma(r)\}$:

Essentially Cyclic rule: There exists $T \ge m$ satisfying $\{1,\ldots,m\} \subseteq \{\sigma(r),\ldots,\sigma(r+T-1)\}$ for all integer $r \ge 1$.

[For example, if we choose $\sigma(km+j) = j$ (for $k = 0,1,\ldots$ and $j = 1,\ldots,m$), we obtain the cyclic relaxation method.] For each $i \in \{1,\ldots,m\}$, denote

$$R_i = \{\, r \ge 1 \mid \sigma(r) = i \,\}.$$

We have the following convergence results regarding the GS-NPPD algorithm (cf. Propositions 1 and 2):

Proposition 3 Let $\{x^r = (x_1^r,\ldots,x_m^r)\}$ be a sequence of iterates generated by the GS-NPPD algorithm under the Essentially Cyclic rule. Then each limit point of $\{x^r\}$ is an optimal solution of (2.1). If furthermore ∇f is Lipschitz continuous on X and each W_i is uniformly continuous on $X_i \times X$, then $[x^r - \nabla f(x^r)]_X^+ - x^r \to 0$.

Proof: For each r, let $y_{\sigma(r)}^r$ be given by (3.1) and let $y_j^r = x_j^r$ for all $j \ne \sigma(r)$. Let $y^r = (y_1^r,\ldots,y_m^r)$ and let $\{x^r\}_{r\in R}$ be a subsequence of $\{x^r\}$ converging to some x^∞. Further passing into a subsequence if necessary, we will assume that $(\sigma(r),\ldots,\sigma(r+T-1))$ is the same for all $r \in R$, say equals $(\sigma_0,\ldots,\sigma_{T-1})$. By an argument analogous to that used in the proof of Proposition 1, we obtain $\{y^r\}_{r\in R} \to x^\infty$. Since $\|x^{r+1}-x^r\| \le \bar{\theta}\|y^r - x^r\|$, this implies that $\{x^{r+1}-x^r\}_{r\in R} \to 0$ and hence $\{x^{r+1}\}_{r\in R} \to x^\infty$. Proceeding in this way, we obtain that $\{x^{r+j}\}_{r\in R} \to x^\infty$ and $\{y^{r+j}\}_{r\in R} \to x^\infty$, for every $j = 0,1,\ldots,T-1$. Since (cf. (3.1))

$$\langle W_{\sigma_j}(y_{\sigma_j}^{r+j},x^{r+j}) - W_{\sigma_j}(x_{\sigma_j}^{r+j},x^{r+j}) + \nabla_{\sigma_j} f(x^{r+j}), y_{\sigma_j} - y_{\sigma_j}^{r+j}\rangle \ \ge\ 0, \quad \forall\, y_{\sigma_j} \in X_{\sigma_j},$$

for all $r \in R$, and W_{σ_j} is continuous on $X_{\sigma_j} \times X$, for $j = 0, 1, \ldots, T-1$, we obtain that

$$\langle \nabla_{\sigma_j} f(x^\infty), y_{\sigma_j} - x_{\sigma_j}^\infty \rangle \ \ge\ 0, \quad \forall\, y_{\sigma_j} \in X_{\sigma_j}, \quad \forall\, j = 0, 1, \ldots, T-1.$$

Since $X = X_1 \times \ldots \times X_m$ and (cf. Essentially Cyclic rule) $\{\sigma_0,\ldots,\sigma_{T-1}\}$ contains $\{1,\ldots,m\}$, this implies $\langle \nabla f(x^\infty), x - x^\infty \rangle \ \ge\ 0$, for all $x \in X$, and hence x^∞ is an optimal solution of (2.1).

Now, suppose that furthermore ∇f is Lipschitz continuous on X and each W_i is uniformly continuous on $X_i \times X$. By an argument analogous to that used in the proof of Proposition 2, we have that $y^r - x^r \to 0$ and, for each $i \in \{1,\ldots,m\}$,

$$\|[x_i^r - \nabla_i f(x^r)]_{X_i}^+ - x_i^r\| \leq \|W_i(y_i^r,x^r) - W_i(x_i^r,x^r)\| + 2\|y_i^r - x_i^r\|, \quad \forall\, r \in R_i.$$

Now, fix any $i \in \{1,\ldots,m\}$ and, for each $r \geq 1$, let $\tau(r)$ denote the smallest integer greater than or equal to r such that $\sigma(\tau(r)) = i$. Since W_i is uniformly continuous on $X_i \times X$, the above inequality and the fact $y^r - x^r \to 0$ implies that

$$[x_i^{\tau(r)} - \nabla_i f(x^{\tau(r)})]_{X_i}^+ - x_i^{\tau(r)} \to 0. \tag{3.2}$$

Then from the triangle inequality and the fact that $x_i^h = x_i^r$ for $h = r, r+1, \ldots, \tau(r)$, for all r, we obtain that, for all r,

$$\begin{aligned}
\|[x_i^r - \nabla_i f(x^r)]_{X_i}^+ - x_i^r\| &\leq \sum_{h=r}^{\tau(r)-1} \|[x_i^r - \nabla_i f(x^h)]_{X_i}^+ - [x_i^r - \nabla_i f(x^{h+1})]_{X_i}^+\| \\
&\quad + \|[x_i^{\tau(r)} - \nabla_i f(x^{\tau(r)})]_{X_i}^+ - x_i^{\tau(r)}\| \\
&\leq \sum_{h=r}^{\tau(r)-1} \|\nabla_i f(x^h) - \nabla_i f(x^{h+1})\| \\
&\quad + \|[x_i^{\tau(r)} - \nabla_i f(x^{\tau(r)})]_{X_i}^+ - x_i^{\tau(r)}\| \\
&\leq \sum_{h=r}^{\tau(r)-1} \lambda \cdot \|x^h - x^{h+1}\|^{\eta} + \|[x_i^{\tau(r)} - \nabla_i f(x^{\tau(r)})]_{X_i}^+ - x_i^{\tau(r)}\|,
\end{aligned}$$

where the second inequality follows from the fact that the projection mapping $[\cdot]_{X_i}^+$ is nonexpansive and the third inequality follows from the Lipschitz continuity of ∇f (λ and η are some positive scalars). Since (cf. Essentially Cyclic rule) $\tau(r) - r \leq T$ for all r, this, together with (3.2) and the fact that $x^{r+1} - x^r \to 0$ (since $\|x^{r+1} - x^r\| \leq \bar{\theta}\|y^r - x^r\|$), implies that $[x_i^r - \nabla_i f(x^r)]_{X_i}^+ - x_i^r \to 0$. Since the choice of i was arbitray, this holds for all $i \in \{1,\ldots,m\}$. Since $X = X_1 \times \ldots \times X_m$, this in turn implies that $[x^r - \nabla f(x^r)]_X^+ - x^r \to 0$.
Q.E.D.

If f has bounded level sets on X and is strictly convex in each x_i, then we can also choose $\bar{\theta} = \infty$ in the GS-NPPD algorithm and Proposition 3 would still hold (it can be shown, using the above assumption, that $x^{r+1} - x^r \to 0$). If X is bounded, then the strong monotonicity assumption on the W_i's can be weakened much as in Proposition 1′. We state this result below. Its proof is analogous to that of Proposition 1′ and is omitted.

Proposition 3′ Assume that X is bounded and let $\{x^r = (x_1^r,\ldots,x_m^r)\}$ be a sequence of iterates generated by the GS-NPPD algorithm under the Essentially Cyclic rule. Then, even if $W_i(\cdot,x)$ is only strictly monotone (instead of strongly monotone) for each $x\in X$ and each i, every limit point of $\{x^r\}$ is an optimal solution of (2.1).

4. **Dual Application**

Consider the convex program

$$\begin{array}{ll} \text{Minimize} & \phi(u) \\ \text{subject to} & Eu \ \geq \ b, \end{array} \tag{4.1}$$

where $\phi:\Re^p\to(-\infty,\infty]$ is a strongly convex function, E is an $n\times p$ matrix, and b is an n-vector. This problem has applications in entropy maximization, linear programming, network programming, and the solution of symmetric linear complementarity problems (see [Tse88b] as well as Examples 3 and 7 in §5). Our results also extend in a straightforward manner to problems with both linear equality and inequality constraints, but for simplicity we will not treat this more general case here. We make the following standing assumptions:

Assumption C:

(a) The function ϕ is closed strongly convex (not necessarily differentiable) and continuous in dom(ϕ).

(b) Dom(ϕ) is the intersection of two convex sets P and Q such that cl(P) is a polyhedral set and $P\cap \mathrm{ri}(Q)\cap\{\, u \mid Eu \geq b\} \neq \varnothing$.

Assumption C (b) is a constraint qualification condition which also implies that (4.1) is feasible. This, together with the fact that ϕ has bounded level sets (since ϕ is strongly convex), implies that (4.1) has an optimal solution which, by the strict convexity of f, is unique. [An example of a convex function ϕ for which cl(dom(ϕ)) is a polyhedral set (but not necessarily dom(ϕ)) is when ϕ is separable.]

By assigning a Lagrange multiplier vector x to the constraints $Eu \geq b$, we obtain the following dual program

$$\begin{array}{ll} \text{Minimize} & f(x) \\ \text{subject to} & x \geq 0, \end{array} \tag{4.2}$$

where $f:\Re^n \to \Re$ is the dual functional given by

$$f(x) \;=\; \max_u\{ \langle x,Eu-b\rangle - \phi(u) \} \;=\; \phi^*(E^Tx) - \langle x,b\rangle.$$

[f is real-valued because ϕ is strongly convex, so that ϕ is co-finite [Roc70, pp. 116].] Since ϕ is strictly convex, f is convex and differentiable. Furthermore, strong duality holds for (4.1) and (4.2), i.e. the optimal value in (4.1) equals the negative of the optimal value in (4.2). [To see this, note that the set $\{ (u,w,z) \mid Eu \geq w, \phi(u) \leq z \}$ is closed. Hence the <u>convex bifunction associated with (4.1)</u> [Roc70, pp. 293] is closed. Since the optimal solution set for (4.1) is bounded, Theorem 30.4 in [Roc70] states that strong duality holds.]

The problem (4.2) is clearly a special case of (2.1) and (cf. Assumption C) Assumption A is satisfied. Furthermore, the constraint set is the Cartesian product of closed intervals. Hence we can apply either the NPPD algorithm or the GS-NPPD algorithm to solve this problem. The resulting methods have characteristics very similar to those of the method of multipliers and the dual descent methods (see §5). Because the level sets of f are not necessarily bounded, these methods are not guaranteed to find an optimal solution of (4.2). [In fact (4.2) may not even have an optimal solution.] On the other hand, we show below that these methods are guaranteed to find the unique optimal solution of (4.1). To show this, we first need the following technical lemma:

Lemma 1 Let $h:\Re^p \to (-\infty,\infty]$ be any closed convex function that is continuous in S = dom(h). Then the following hold:

(a) For any $u \in S$, there exists a positive scalar ε such that $S \cap B(u,\varepsilon)$ is closed, where $B(u,\varepsilon)$ denotes the closed ball around u with radius ε.

(b) For any $u \in S$, any z such that $u + z \in S$, and any sequences $\{u^k\} \to u$ and $\{z^k\} \to z$ such that $u^k \in S$, $u^k + z^k \in S$ for all k, we have

$$\lim_{k\to\infty} \sup\{h'(u^k;z^k)\} \leq h'(u;z).$$

(c) If h is furthermore co-finite, then for any $u \in S$ and any sequence $\{u^k\} \in S$ such that $\{h(u^k) + h'(u^k;u-u^k)\}$ is bounded from below, we have that both $\{u^k\}$ and $\{h(u^k)\}$ are bounded, and every limit point of $\{u^k\}$ is in S.

Proof: We prove (a) only. Parts (b) and (c) follow from, respectively, the proof of Lemmas 3 and 2 in [TsB87]. Let $Y = cl(S)\backslash S$. It then suffices to show that $Y \cap B(u,\varepsilon) = \varnothing$ for some positive ε. Suppose the contrary. Then there exists a sequence of points $\{y^1,y^2,\ldots\}$ in Y converging to u. Consider a fixed k. Since $y^k \in Y$, there exists a sequence of points $\{w^{k,1},w^{k,2},\ldots\}$ in S converging to y^k. Since h is closed, it must be that $h(w^{k,i}) \to \infty$ as $i \to \infty$. Therefore there exists integer m(k) for which $\|w^{k,m(k)}-y^k\| < 1/k$ and $w^{k,m(k)} > k$. Then $\{w^{k,m(k)}\}_{k=1,2,\ldots}$ is a sequence of points in S converging to u for which $h(w^{k,m(k)}) \to \infty$, contradicting the continuity of h on S since $h(u) < \infty$.
Q.E.D.

By combining the above lemma with Propositions 2 and 3, we obtain the main result of this section:

Proposition 4 If $\{x^r\}$ is a sequence of iterates generated by the NPPD algorithm applied to solving (4.2) and W is uniformly continuous on $\Re^n_+ \times \Re^n_+$, then $\{\nabla\phi^*(E^T x^r)\}$ converges to the optimal solution of (4.1). The same conclusion holds if $\{x^r\}$ is a sequence of iterates generated by the GS-NPPD algorithm under the Essentially Cyclic rule and each W_i is uniformly continuous on $\Re^{n_i}_+ \times \Re^n_+$ $(i = 1, \ldots, m)$.

Proof: Since ϕ is strongly convex, ∇f is Lipschitz continuous. Hence by Propositions 2 and 3,

$$[x^r - \nabla f(x^r)]^+ - x^r \to 0, \qquad (4.3)$$

where $[\cdot]^+$ denotes the orthogonal projection onto $\Re^n_+$ and $\{x^r\}$ is a sequence of iterates generated by either the NPPD algorithm or the GS-NPPD algorithm under the Essentially Cyclic rule. To simplify the notation, let $u^r = \nabla\phi^*(E^Tx^r)$, let $\bar{u}$ be the optimal solution of (4.1), and let U be the constraint set for (4.1), i.e. $U = \{u \mid Eu \geq b\}$. Then (cf. [Roc70, Theorem 23.5]) $f(x^r) = \langle x^r, Eu^r-b\rangle - \phi(u^r)$ and $Eu^r-b = \nabla f(x^r)$ for all r; hence, for all r,

$$\begin{aligned} f(x^r) &\geq \langle x^r, Eu^r-b\rangle - \phi(u^r) - \langle x^r, E\bar{u}-b\rangle \\ &= -\phi(u^r) - \langle E^Tx^r, \bar{u}-u^r\rangle \\ &\geq -\phi(u^r) - \phi'(u^r;\bar{u}-u^r), \end{aligned}$$

where the first inequality follows from the fact that $\bar{u}$ is feasible for (4.1) and $x^r \geq 0$ for all r; the second inequality follows from the fact that $E^Tx^r \in \partial\phi(u^r)$. Since $\{f(x^r)\}$ is bounded from above by $f(x^0)$ and ϕ is closed, co-finite and continuous in dom(ϕ), the above inequality, together with Lemma 1 (c), implies that the sequence $\{u^r\}$ is bounded and everyone of its limit points is in dom(ϕ). Let u^∞ be any limit point of $\{u^r\}$. Since (cf. (4.3)) $[x^r-Eu^r+b]^+ - x^r \to 0$ and $x^r \geq 0$ for all r, we have, upon passing into the limit, that $Eu^\infty \geq b$. Hence $u^\infty \in \text{dom}(\phi)\cap U$.

We claim that $u^\infty = \bar{u}$. To see this, suppose that $u^\infty \neq \bar{u}$ and let y be an element of $P\cap\text{ri}(Q)\cap U$. Fix any $\lambda \in (0,1)$ and denote $y(\lambda) = \lambda y+(1-\lambda)\bar{u}$. Then $y(\lambda) \in P\cap\text{ri}(Q)\cap U$ and $y(\lambda) \neq u^\infty$. Let $\{u^r\}_{r\in R}$ be a subsequence of $\{u^r\}$ converging to u^∞. By Lemma 1 (a), there exists an $\varepsilon > 0$ such that $\text{dom}(\phi)\cap B(u^\infty,\varepsilon) = \text{cl}(P)\cap Q\cap B(u^\infty,\varepsilon)$. Since cl(P) is a polyhedral set and $y(\lambda)-u^\infty$ belongs to the tangent cone of cl(P) at u^∞, this implies that, for any $\delta \in (0,\varepsilon)$,

$$u^r + \delta z \in P, \qquad \forall\, r \in R \text{ sufficiently large,} \tag{4.4}$$

where $z = (y(\lambda)-u^\infty)/\|y(\lambda)-u^\infty\|$. On the other hand, since $y(\lambda) \in \text{ri}(Q)$, $u^r \in Q$ for all r, and $\{u^r\}_{r\in R} \to u^\infty$, we have

$$u^r + \delta z \in Q, \qquad \forall\, r \in R \text{ sufficiently large.} \tag{4.5}$$

Since $E^Tx^r \in \partial\phi(u^r)$ for all r, $\phi'(u^r;z) \geq \langle x^r, Ez\rangle$ for all r. Since $u^\infty + \delta z \in \text{dom}(\phi)$ and ϕ is continuous in dom(ϕ), this, together with (4.4)-(4.5) and Lemma 1 (b), implies that

$$\phi'(u^\infty;z) \geq \lim_{r\to\infty, r\in R} \inf \langle x^r, Ez\rangle.$$

Now $[x^r - Eu^r + b]^+ - x^r \to 0$ implies that $[Eu^\infty]_i = b_i$ for all i such that $\{x_i^r\}_{r\in R} \not\to 0$. Since $Ey(\lambda) \geq b$, this in turn implies that $[Ez]_i \geq 0$ for all i such that $\{x_i^r\}_{r\in R} \not\to 0$, so that

$$\lim_{r\to\infty, r\in R} \inf \langle x^r, Ez\rangle \geq 0.$$

Hence $\phi'(u^\infty;z) \geq 0$ and therefore $\phi(u^\infty) \leq \phi(y(\lambda))$. Since the choice of $\lambda \in (0,1)$ was arbitrary, by taking λ arbitrarily small (and using the continuity of ϕ within $\mathrm{dom}(\phi)$), we obtain that $\phi(u^\infty) \leq \phi(\bar{u})$. Since $u^\infty \in U$, u^∞ is an optimal solution of (4.1). But since (4.1) has a unique optimal solution $\bar{u}$, it holds that $u^\infty = \bar{u}$. Q.E.D.

5. Applications

Below we give some applications of the NPPD algorithm and the GS-NPPD algorithm and show that they are closely related to a number of existing algorithms.

Example 1 (Gabay's Algorithm) Consider the special case of problem (2.1) where f is the sum of two continuously differentiable functions $g:X \rightarrow \Re$ and $h:X \rightarrow \Re$. If we apply the NPPD algorithm to solve this problem with $W(x,y) = \nabla h(x) + x/c$ $(c > 0)$, the y^r's generated from solving the variational inequalities (2.2) satisfy

$$y^r = \operatorname{argmin}_{x \in X} \{ h(x) + \|x - x^r\|^2/2c + \langle \nabla g(x^r), x \rangle \},$$

or equivalently,

$$y^r = [I + c(\nabla h + \Gamma)]^{-1} [I - c\nabla g](x^r),$$

where $\Gamma(\cdot)$ is the subdifferential of the indicator function for X. Hence the NPPD algorithm without the line search is exactly the splitting algorithm proposed by Gabay [Gab83]. In contrast to Gabay's algorithm, neither h nor g has to be convex here (as long as the function $x \rightarrow \nabla h(x) + x/c$ is strongly monotone) and c can be any positive scalar, but an extra line search is needed at each iteration.

Example 2 (Gradient Projection and Proximal Minimization) Consider applying the NPPD algorithm to solve the problem (2.1) with $W(x,y) = \|x\|^2/(2c)$ $(c > 0)$. Then the y^r's generated from solving the variational inequalities (2.2) satisfy

$$y^r = \operatorname{argmin}_{x \in X} \{ \|x - x^r + c\nabla f(x^r)\|^2 \}.$$

Hence the NPPD algorithm without the line search is exactly the gradient projection algorithm [Gol64], [LeP66]. If we let $W(x,y) = \nabla f(x) + x/c$ instead, then y^r is given by

$$y^r = \operatorname{argmin}_{x \in X} \{ f(x) + \|x - x^r\|^2/2c \},$$

so that the NPPD algorithm without the line search is exactly the proximal minimization algorithm [Mar70], [Roc76a].

Example 3 (GP-SOR Algorithm) Consider the special case of (2.1) where $X = \Re^n_+$ and $f(x) = \langle x,Mx\rangle + \langle b,x\rangle$, where M is an n×n symmetric positive semidefinite matrix and b is an n-vector. This is commonly known as the symmetric linear complementarity problem. Let us apply the NPPD algorithm to solve this problem with $W(x,y) = Qx$, where Q is any n×n symmetric positive definite matrix. Then the y^r generated from solving the variational inequality (2.2) is given by

$$y^r = [\, y^r - (Qy^r + (M-Q)x^r + b)\,]^+.$$

where $[\cdot]^+$ denotes the orthogonal projection onto $\Re^n_+$, or equivalently,

$$y^r = [\, y^r - \omega E(Qy^r + (M-Q)x^r + b)\,]^+,$$

where E is any n×n diagonal matrix with positive diagonal entries and ω is any positive scalar. If we let $K = Q - (\omega E)^{-1}$, then the above equation becomes:

$$y^r = [\, x^r - \omega E(Mx^r + K(y^r - x^r) + b)\,]^+.$$

Hence the NPPD algorithm in this case is exactly the GP-SOR algorithm of Mangasarian and De Leone [MaD88]. The condition that $Q = (\omega E)^{-1} + K$ be positive definite is identical to the condition (3.3) in [MaD88]. Some choices for Q that allows y^r to be computed fairly easily are (cf. [MaD88, Corollary 3.3]) (i) Q is diagonal or tridiagonal, and (ii) $E = D^{-1}$ and $Q = \omega^{-1}D + L$, where D (L) is the diagonal (strict lower triangular) part of M and $\omega \in (0,2)$ (assuming that M has positive diagonal entries). [Note that f has bounded level sets on $\Re^n_+$ if and only if there exists $z \in \Re^n$ such that $Mz + b > 0$.]

The convergence of the sequence $\{x^r\}$ generated by the GP-SOR algorithm remains an open question. However, we can prove a slightly weaker (but still very useful) result by using Proposition 4:

Proposition 5 Consider any $n \times p$ matrix A such that $M = AA^T$ and any p-vector w. If $\{x^r\}$ is any sequence generated by the GP-SOR algorithm, then the sequence $\{A^T x^r - w\}$ converges to the unique optimal solution of the convex quadratic program

$$\begin{array}{ll} \text{Minimize} & \|u\|^2/2 + \langle w,u\rangle \\ \text{subject to} & Au \geq -Aw - b. \end{array} \tag{5.1}$$

Proof: The problem (5.1) is a special case of (4.1). By assigning a nonnegative Lagrange multiplier vector x to the constraints $Au \geq -Aw-b$, we obtain the dual of (5.1) to be exactly the symmetric linear complementarity problem

$$\begin{array}{ll} \text{Minimize} & \langle x,Mx\rangle + \langle b,x\rangle \\ \text{subject to} & x \geq 0. \end{array}$$

Since this problem by assumption has an optimal solution, its dual (5.1) must be feasible. Hence (5.1) satisfies Assumption C. By Proposition 4, the sequence $\{A^T x^r - w\}$ converges to the unique optimal solution of (5.1). Q.E.D.

As a corollary, we have that $\{Mx^r\}$ converges to a unique limit point and that the linear programming algorithm in [MaD88, §5] is convergent. Alternatively, we can apply the GS-NPPD algorithm to solve the above problem, say with $W_i(x_i,y) = Q_{ii}x_i$, where Q_{ii} is some $n_i \times n_i$ symmetric positive definite matrix. This gives a GS-NPPD method that is potentially faster (though less parallelizable) than its Jacobi cousin, the GP-SOR algorithm. Furthermore, if M has a block diagonal structure, then we can partition the components of x such that each x_i corresponds to a block. Convergence of this method also follows from Proposition 4.

Example 4 (Coordinate Descent) Consider the special case of (2.1) where $X = X_1 \times \ldots \times X_m$, for some closed convex sets $X_1, \ldots, X_m$ (cf. Assumption B). Furthermore let us assume that f has bounded level sets on X and that $f(x_1,\ldots,x_m)$ is strictly convex in each $x_i \in X_i$ (with the other x_j's fixed). Let us apply the GS-NPPD algorithm to solve this problem with W_i chosen to be

$$W_i(x_i,y) = \nabla_i f(y_1,\ldots,y_{i-1},x_i,y_{i+1},\ldots,y_m), \qquad \forall\, i.$$

Then it is easily seen that, for each r, y_i^r given by (3.1) minimizes $f(x_1^r,\ldots,x_{i-1}^r,x_i,x_{i+1}^r,\ldots,x_m^r)$ over all $x_i \in X_i$. Hence if we choose $\bar{\theta} = +\infty$ in the GS-NPPD algorithm, then $x_i^{r+1} = y_i^r$ for all r and the algorithm reduces to the classical coordinate descent method [D'Es59], [Lue84], [Pow73], [SaS73], [Zan69]. Because the level sets of f are bounded, the GS-NPPD algorithm is effectively operating on a compact subset of $X_1 \times \ldots \times X_m$, and it follows from Proposition 3′ that it converges (see [BeT89, §3.3.5], [Tse88c] for related results).

Example 5 (Traffic Assignment) Consider a directed transportation network consisting of p nodes and n arcs. On this network, a total of m commodities are to be sent from certain origins to certain destinations. Associated with the jth arc is a scalar cost function f_j. A total of θ units of commodities sent on the jth arc incurs a cost of $f_j(\theta)$. The objective is to determine the amount of each commodity to send on each arc in order to minimize the sum of the arc costs while meeting all of the demands. This problem, known as the traffic assignment problem [AaM81], [BeG82], [ChM88], [Daf80], can be formulated as the following nonlinear multicommodity problem:

$$\begin{array}{lll} \text{Minimize} & \sum_{j=1}^{n} f_j(x_{1j}+x_{2j}+\ldots+x_{mj}) & \\ \text{subject to} & Ax_1 = b_1, & \\ & \quad Ax_2 = b_2, & \\ & \qquad \vdots & \\ & \qquad\quad Ax_m = b_m, & \\ & 0 \le x_{ij} \le u_{ij}, \;\; i = 1,2,\ldots,m, \;\; j = 1,2,\ldots,n, & \end{array} \tag{5.2}$$

where A is the p×n node-arc incidence matrix for the network, $u_{ij} \ge 0$ is the capacity of the jth arc for the ith commodity, x_{ij} is the amount of the ith commodity sent on the jth arc, x_i is the n-vector $(x_{i1},x_{i2},\ldots,x_{in})$, and $f_j:[0,\sum_i u_{ij}] \to \Re$ is a convex differentiable function. Let X denote the Cartesian product $[0,u_{11}] \times [0,u_{12}] \times \ldots \times [0,u_{mn}]$. We assume that (5.2) has a feasible solution, which, in view of the compactness of X, implies that (5.2) has an optimal solution.

The problem (5.2) is clearly a special case of (2.1) and therefore we can use either the NPPD algorithm or the GS-NPPD algorithm to solve (5.2). If we use the NPPD algorithm, a reasonable choice for $W:X\times X\rightarrow\Re^{mn}$ that uses information about f is the function given by (cf. (2.4))

$$W(x,y) = (\ldots, \nabla f_j(x_{ij} + \textstyle\sum_{k\neq i} y_{kj}) + \rho_{ij}(x_{ij}-y_{ij}) , \ldots)_{i=1,\ldots,m;\ j=1,\ldots,n},$$

where each ρ_{ij} is a positive scalar. [If f_j is furthermore strictly convex, then (cf. Proposition 1′) $\rho_{ij} = 0$ is also permissible.] With this choice of W, each iteration involves the solution of m separate single commodity flow problems and, like the Chen-Meyer algorithm [ChM88], these problems can be solved in parallel on a MIMD machine. If we use the GS-NPPD algorithm, a reasonable choice for X_i is $X_i = [0,u_{i1}]\times\ldots\times[0,u_{in}]$ and for $W_i:X_i\times X\rightarrow\Re^{n+mn}$ is the function

$$W_i(x_i,y) = (\ldots, \nabla f_j(x_{ij} + \textstyle\sum_{k\neq i} y_{kj}) + \rho_{ij}(x_{ij}-y_{ij}) , \ldots)_{j=1,\ldots,n},$$

where each ρ_{ij} is a positive scalar. [If each f_j is furthermore strictly convex, then $\rho_{ij} = 0$ is also permissible.] In either case, each single commodity flow problem that we solve has a separable strictly convex cost and can be solved by one of a number of methods [BHT87], [FeM88], [DeK81], [Mil63], [Roc83].

Example 6 (Dual Gradient Method) Consider the convex program (4.1) (under Assumption C). Suppose that we apply the NPPD algorithm to solve this program with $W(x,y) = Hx$, where H is an $n\times n$ symmetric positive definite matrix. Since W is uniformly continuous, the resulting method converges in the sense of Proposition 4. For quadratic cost problems with equality constraints, this reduces to the dual gradient method proposed in [LiP87, Theorem 4.4.1] and in [Pan86, §6]. In practice, this method can be implemented as follows: At the rth iteration, first compute u^r to be the unique minimizer of the function $\phi(u) - \langle x^r,Eu\rangle$ over all u. Then update the Lagrange multiplier vector by

$$x^{r+1} = x^r + \theta^r d^r,$$

where $d^r = H^{-1}(b-Eu^r)$ and θ^r is the Kuhn-Tucker vector for the knapsack problem $\min\{\phi(u) - \langle x^r,Eu\rangle \mid \langle d^r,b-Eu\rangle = 0\}$ if $d^r \geq 0$; otherwise θ^r is $\bar{\theta}^r$ plus the Kuhn-Tucker vector for the knapsack problem $\min\{\phi(u) - \langle x^r + \bar{\theta}^r d^r,Eu\rangle \mid \langle d^r,b-Eu\rangle \leq 0\}$, where $\bar{\theta}^r$ is the largest θ for which $x^r + \theta d^r \geq 0$.

Example 7 (Dual Splitting Method) Consider the convex program (4.1) (under Assumption C), and let us further assume that the cost function ϕ has the following separable form

$$\phi(w,z) = \eta(w) + \psi(z),$$

where $\eta:\Re^{m_1}\to(-\infty,\infty]$ and $\psi:\Re^{m_2}\to(-\infty,\infty]$ are closed strongly convex functions ($m_1+m_2 = m$). We show below that, by choosing the function W in the NPPD algorithm appropriately, we obtain a decomposition algorithm for solving this separable program.

Let us partition E into E = [A B] corresponding to w and z. Then the dual of this convex program (cf. (4.2)) can be written as

$$\begin{array}{ll} \text{Minimize} & h(x) + g(x) \\ \text{subject to} & x \geq 0, \end{array}$$

where $h:\Re^n\to\Re$ and $g:\Re^n\to\Re$ are dual functionals given by

$$\begin{aligned} h(x) &= \eta^*(A^Tx) - \langle x,b\rangle, \\ g(x) &= \psi^*(B^Tx). \end{aligned}$$

Let us apply the NPPD algorithm to solve this dual problem with $W(x,y) = \nabla h(x) + x/c$ ($c > 0$) for all r, and let $\{x^r\}$, $\{y^r\}$ denote the sequence of iterates thus generated. Then we have (cf. Example 1)

$$y^r = [\, I + c(\nabla h + \Gamma)\,]^{-1}[\, I - c\nabla g\,](x^r),$$

where Γ denotes the subdifferential of the indicator function for $\Re^n_+$, or equivalently,

$$y^r = [\, x^r + c(b - Aw^r - Bz^r)\,]^+,$$

where $[\cdot]^+$ denotes the orthogonal projection onto $\Re^n_+$ and z^r and w^r are given by

$$z^r = \operatorname{argmin}_z\{\, \psi(z) - \langle x^r,Bz\rangle\, \},$$

$$w^r = \text{argmin}_w\{\ \eta(w) + \|[\ x^r + c(b - Aw - Bz^r)\ ]^+\|^2/2c\ \}.$$

Hence y^r can be obtained by solving two (explicitly) unconstrained problems in z and in w respectively. In fact, the above computation of y^r is equivalent to an iteration of the alternating minimization algorithm [Tse88a], but has the additional advantage that c is not upper bounded by the curvature of ψ. On the other hand, we require that η be strongly convex and that an additional line search be made at each iteration. [The line search can be performed by solving a knapsack problem analogous to that described in Example 6.]

6. Extensions

There are a number of directions in which our results can be extended. For example, we can use the more general function $W(\cdot,x^r,x^{r-1},\ldots,x^{r-d})$ $(d \geq 1)$ in the NPPD algorithm instead of $W(\cdot,x^r)$. This would allow more of the past history to be used. Also we can allow W to change dynamically, i.e. replace W by some strongly monotone function W^r at the rth iteration. It can be shown that Proposition 1 still holds provided that in addition $\{W^1, W^2, \ldots\}$ is a family of pointwise equi-continuous functions in the sense that, for any $\bar{x} \in X$ and any $\varepsilon > 0$, there exists $\delta > 0$ such that $\|W^r(y,x)-W^r(x,x)\| \leq \varepsilon$ for all r and all x, $y \in X$ such that $\|\bar{x}-x\| \leq \delta$, $\|\bar{x}-y\| \leq \delta$. Similarly, it can be shown that Proposition 2 still holds provided that $\{W^1, W^2, \ldots\}$ is a family of equi-continuous functions (in addition to being strongly monotone) in the sense that, for any $\varepsilon > 0$, there exists $\delta > 0$ such that $\|W^r(y,x)-W^r(x,x)\| \leq \varepsilon$ for all r and all x, $y \in X$ such that $\|x-y\| \leq \delta$. [For example, we can in Example 6 permit the matrix H to change with each iteration, provided that the eigenvalue of H remains bounded.] Analogous generalizations also hold for the GS-NPPD algorithm.

A generalization of the GS-NPPD algorithm is to choose a finite collection of nonempty subsets $M_j \subseteq \{1,\ldots,m\}$ $(j = 1,\ldots,K)$ such that their union $M_1 \cup \ldots \cup M_K$ equals $\{1,\ldots,m\}$. [The M_j's do not have to be disjoint.] At each iteration, we choose an index $j \in \{1,\ldots,K\}$ and solve a variational inequality analogous to (3.1) that involves $\{x_i\}_{i \in M_j}$ as the variables. Under the assumption that there exists $T' > 0$ such that all elements of $\{1,\ldots,K\}$ are chosen during every T' consecutive iterations, it can be shown that a conclusion analogous to that of Proposition 3 holds.

References

[AaM81] Aashtiani, H.Z. and Magnanti, T.L., "Equilibria on a Congested Transportation Network," *SIAM J. Algebraic and Discrete Math.*, 2 (1981), 213-226.

[Ber82] Bertsekas, D.P., *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, NY (1982).

[BeG82] Bertsekas, D.P. and Gafni, E.M., "Projection Methods for Variational Inequalities with Application to the Traffic Assignment Problem," *Math. Prog. Study*, 17 (1982), 139-159.

[BHT87] Bertsekas, D.P., Hosein, P.A. and Tseng, P., "Relaxation methods for network flow problems with convex arc costs," *SIAM J. Control and Optimization*, 25 (1987), 1219-1243.

[BeT89] Bertsekas, D.P. and Tsitsiklis, J.N., *Parallel and Distributed Computation: Numerical Methods*, Prentice-Hall, Englewood Cliffs, NJ (1989).

[ChM88] Chen, R.J. and Meyer, R.R., "Parallel Optimization for Traffic Assignment," *Math. Prog.*, 42 (1988), 327-345.

[Daf80] Dafermos, S.C., "Traffic Equilibrium and Variational Inequalities," *Transport. Sci.*, 14 (1980), 42-54.

[D'Es59] D'Esopo, D.A., "A Convex Programming Procedure," *Naval Research Logistics Quarterly*, 6 (1959), 33-42.

[DeK81] Dembo, R.S. and Klincewicz, "A Scaled Reduced Gradient Algorithm for Network Flow Problems with Convex Separable Costs," *Math. Prog. Study*, 15 (1981), 125-147.

[FeM88] Feijoo, B. and Meyer, R.R., "Piecewise-Linear Approximation Methods for Nonseparable Convex Programming," *Management Science*, 34 (1988), 411-419.

[Gab83] Gabay, D., "Applications of the Method of Multipliers to Variational Inequalities," in M. Fortin and R. Glowinski, eds., *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*, North-Holland, Amsterdam (1983).

[Gol64] Goldstein, A.A., "Convex Programming in Hilbert Space," *Bull. Amer. Math. Soc.*, 70 (1964), 709-710.

[Kor76] Korpelevich, G.M., "The Extragradient Method for Finding Saddle Points and Other Problems," *Ekon. i Mat. Metody*, 12 (1976), 747-756; Engl. transl. in *Matekon.*

[LeP66] Levitin, E.S. and Poljak, B.T., "Constrained Minimization Methods," *Z. Vycisl. Mat. i mat. Fiz.*, 6, 787-823; Engl. transl. in *C.M.M.P.*, 6 (1966), 1-50.

[LiP87] Lin, Y.Y. and Pang, J.-S., "Iterative Methods for Large Convex Quadratic Programs: A Survey," *SIAM J. Control and Optimization*, 25 (1987), 383-411.

[LiM79] Lions, P.L. and Mercier, B., "Splitting Algorithms for the Sum of Two Nonlinear Operators," *SIAM J. Numerical Analysis*, 16 (1979), 964-979.

[Lue84] Luenberger, D.G., *Linear and Nonlinear Programming*, Addison-Wesley, Reading, MA (1984).

[Luq86] Luque, J., "The Nonlinear Proximal Point Algorithm," LIDS Report P-1598, M.I.T., Cambridge, MA (August 1986).

[MaD88] Mangasarian, O.L. and De Leone, R., "Parallel Gradient Projection Successive Overrelaxation for Symmetric Linear Complementarity Problems," in *Annals of Operations Research 14* (1988), 41-59.

[Mar70] Martinet, B., "Regularisation d'inequations variationnelles par approximations successives," *Rev. Française Inf. Rech. Oper.* (1970), 154-159.

[Mil63] Miller, C.E., "The Simplex Method for Local Separable Programming," in *Recent Advances in Mathematical Programming* (R.L. Graves and P. Wolfe, eds.), McGraw-Hill, New York, NY (1963), 89-100.

[Pan86] Pang, J.-S., "More Results on the Convergence of Iterative Methods for the Symmetric Linear Complementarity Problem," *J. Optimization Theory and Applications*, 49 (1986), 107-134.

[Pow73] Powell, M.J.D., "On Search Directions for Minimization Algorithms," *Math. Prog.*, 4 (1973), 193-201.

[Roc70] Rockafellar, R.T., *Convex Analysis*, Princeton University Press, Princeton, NJ (1970).

[Roc76a] Rockafellar, R. T., "Augmented Lagrangians and Applications for the Proximal Point Algorithm in Convex Programming," *Mathematics of Operations Research*, 1 (1976), 97-116.

[Roc76b] Rockafellar, R. T., "Monotone Operators and the Proximal Point Algorithm," *SIAM J. Control and Optimization*, 14 (1976), 877-898.

[Roc84] Rockafellar R.T., *Network Flows and Monotropic Optimization*, Wiley-Interscience, New York, NY (1984).

[SaS73] Sargent, R.W.H. and Sebastian, D.J., "On the Convergence of Sequential Minimization Algorithms," *J. Optimization Theory and Applications*, 12 (1973), 567-575.

[Tse88a] Tseng, P., "Applications of a Splitting Algorithm to Decomposition in Convex Programming and Variational Inequalities," LIDS Report P-1836, M.I.T., Cambridge, MA (November 1988).

[Tse88b] Tseng, P., "Dual Coordinate Ascent for Problems with Strictly Convex Costs and Linear Constraints: A Unified Approach," LIDS Report P-1792, M.I.T., Cambridge, MA (July 1988); to appear in *SIAM J. Control and Optimization*.

[Tse88c] Tseng, P., "Coordinate Ascent for Maximizing Nondifferentiable Concave Functions," LIDS Report P-1840, M.I.T., Cambridge, MA (December 1988).

[TsB87] Tseng, P. and Bertsekas, D.P., "Relaxation Methods for Problems with Strictly Convex Costs and Linear Inequality Constraints," LIDS Report P-1717, M.I.T., Cambridge, MA (November 1987).

[Zan69] Zangwill, W.I., *Nonlinear Programming*, Prentice-Hall, Englewood Cliffs, NJ (1969).