Convergent Net Weighting Schemes in Hypergraph-based Optimization *

Tony F. Chan[1,2,3], Jason Cong[1], Eric M. Radke[2]
[1]Univ. of California, Los Angeles Computer Science Dept.
[2]Univ. of California, Los Angeles Mathematics Dept.
[3]Hong Kong University of Science and Technology
tonyfchan@ust.hk, cong@cs.ucla.edu, radke@math.ucla.edu

October 3, 2011

1 Introduction

Approaches for solving the timing-driven placement problem have traditionally been either net-based or path-based; see, e.g., [7] for an overview. *Net weighting* methods, which fall in the latter category, have been a popular tool in analytical placers [16, 8, 9] for handling timing-driven placement. They enjoy a number of advantages, including very low computational complexity, high flexibility, and ease of implementation – weighting algorithms can be implemented within the framework of an existing placement tool by a simple modification of the objective function. However, net weighting methods suffer the disadvantage that they are largely *ad-hoc*; to date there has been very little theoretical justification for their use [10]. As a result, a number of very different weighting schemes have been proposed, of which some have been shown to be effective in reducing delay. Our goal is to distill the essential properties of a robust and effective net weighting method.

Among commonly used schemes is the *VPR weighting* [11], a *polynomial* scheme defined on the edges by

$$w_e = \left(1 - \frac{\sigma(e)}{T_p}\right)^{\alpha} \tag{1.1}$$

where $\sigma(e)$ is the slack of edge e, T_p is the max path delay of the previous iterate, and α is a user-defined constant. Another scheme is the *PATH weighting* [10], an *exponential* scheme that considers all paths in a circuit efficiently:

$$w_e = \sum_{\pi \ni e} \alpha^{-\frac{\sigma(\pi)}{T}} \tag{1.2}$$

where $\sigma(\pi)$ is the slack of path π, T is a desired max path delay, and α is again a user-defined constant. A third scheme is the *APlace weighting* [9], a *piecewise polynomial* scheme given by

$$\begin{aligned} w_e &= \sum_{\pi \ni e} f\big(delay(\pi), T_u\big), \\ \text{where } f(d, T_u) &= \begin{cases} \left(\frac{d}{T_u}\right)^{\alpha} - 1 & \text{if } d > T_u \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{1.3}$$

*This research is partially supported by the National Science Foundation under grants CCF-0430077 and CCF-0528583.

Here $T_u = (1-u)T_p$, where u is a constant selected to be 0.1, 0.2, or 0.3, and α is also a user-defined constant.

It has remained an open question under what conditions a net weighting scheme will converge to a timing feasible placement, and what quality can be expected of such a placement. These are addressed in the contributions of this work:

1. We develop a rigorous, generalized framework under which certain net weighting schemes are *guaranteed* to converge to the optimum of the original timing-constrained placement problem, provided the net weighted objective is minimized to a satisfactory degree. We then identify particular net weighting schemes that adhere to this framework.
2. In the case we are able to find global minimizers of unconstrained net weighted objectives, convergence is guaranteed to the global minimizer of the original timing-constrained problem.
3. However, most placers in practice cannot find a global minimizer and instead search for *approximate* local minimizers of the net weighted objectives. In this case, convergence is still guaranteed to a local minimum candidate point of the original timing-constrained problem.
4. We implement convergent weighting schemes in the state-of-the-art placer mPL [3, 12]. When we compare one scheme, a modification of the VPR weighting, to the original method, we find an average 4% delay improvement on the MCNC benchmarks [18]. In every one of the benchmarks tested, delay improvement was superior for the modified weighting.

2 Preliminaries

2.1 Definitions and Problem Formulation

We use the timing-driven placement problem as the most immediate application of this framework, and refer to it throughout the remainder of this paper. However, the framework need not be restricted to placement, as it can be considered in a general hypergraph-based optimization problem in which net weights are applied to handle timing constraints.

Suppose we wish to minimize total wire length of a circuit by determining the pin locations $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, subject to the constraint that the total delay along any path should not exceed an upper bound $T > 0$. The objective we wish to minimize is given by

$$G(\mathbf{x}) = \sum_{\text{nets } i \in \mathcal{N}} h_i(\mathbf{x})$$

where $\mathcal{N}$ is the (finite) set of all nets in the circuit, and $h_i(\mathbf{x})$ is a *continuous, nonnegative* function measuring net $i = \{i_1, i_2, \ldots, i_m\}$ with the property

$$h_i(\mathbf{x}) = 0 \implies x_{i_1} = x_{i_2} = \cdots = x_{i_m} \tag{2.1}$$

For h_i, we could use for example half-perimeter wire length or its log-sum-exp approximation [15, 9],

$$h_i^{\text{lse}}(\mathbf{x}) = \eta \log \sum_{j=1}^{m} \exp(x_{i_j}/\eta) + \eta \log \sum_{j=1}^{m} \exp(-x_{i_j}/\eta)$$

for $\eta > 0$ small. The quadratic wire length approximation, as in e.g. [8], given by

$$h_i^{\text{quad}}(\mathbf{x}) = \sum_{j=1}^{m} \sum_{k=j+1}^{m} h_{i_j,i_k} |x_{i_j} - x_{i_k}|^2$$

is also suitable in this framework.

To measure delay along each edge e in the circuit, we use the *delay function* $d_e(\mathbf{x})$. (For each source-sink pair of pins connected by a net, we consider an edge connecting them. We may also consider internal delay of a node by considering an edge between an input and output pin of the node.) The delay function $d_e(\mathbf{x})$ measuring the delay of edge $e = (s,t)$ is a function of the overall placement $\mathbf{x}$, and is *continuous, nonnegative, convex*, and has the property

$$x_s = x_t \implies d_e(\mathbf{x}) = 0 \tag{2.2}$$

For example, the Euclidean and quadratic delay functions, given by $d_e(\mathbf{x}) = \gamma_e|x_t - x_s|$, and $d_e(\mathbf{x}) = \gamma_e(x_t - x_s)^2$, respectively, are suitable in this framework. For simplicity we have assumed that the placement area is one-dimensional. The analogous formulation for a two- or three-dimensional placement follows in a straightforward manner. Note in particular that the Euclidean distance delay function is suitable in higher dimensions, e.g. $d_e(\mathbf{x}, \mathbf{y}) = \gamma_e\sqrt{(x_t - x_s)^2 + (y_t - y_s)^2}$.

In the interest of brevity, we often use the *edge delay vector* $\mathbf{d}(\mathbf{x})$, which is defined simply as the vector of delays $d_e(\mathbf{x})$ along each edge e in the circuit, i.e. $\mathbf{d}(\mathbf{x}) = [d_1(\mathbf{x}) \ \dots \ d_M(\mathbf{x})]^T$, where M is the number of edges in the circuit.

Further, we consider the *slacks* of the circuit, which are functions of the edge delay vector. For each path π we have the *path slack* σ_π:

$$\sigma_\pi(\mathbf{d}(\mathbf{x})) = \ T - \sum_{\text{edges } e\in\pi} d_e(\mathbf{x}) \tag{2.3}$$

We require the path slack to be nonnegative in our problem formulation.

Also useful in this framework is the notion of *edge slack.* First define the *arrival time* of pin t, $A_t(\mathbf{d}(\mathbf{x}))$, recursively as follows:

$$A_t(\mathbf{d}(\mathbf{x})) = \begin{cases} 0 & \text{if } t \in \mathcal{P}_\mathcal{I} \\ \max_{s\in\text{fanin}(t)}\{A_s(\mathbf{d}(\mathbf{x})) + d_{(s,t)}(\mathbf{x})\} & \text{otherwise} \end{cases} \tag{2.4}$$

Similarly we define the *required arrival time* of pin s, $R_s(\mathbf{d}(\mathbf{x}))$, as follows:

$$R_s(\mathbf{d}(\mathbf{x})) = \begin{cases} T & \text{if } s \in \mathcal{P}_\mathcal{O} \\ \min_{t\in\text{fanout}(s)}\{R_t(\mathbf{d}(\mathbf{x})) - d_{(s,t)}(\mathbf{x})\} & \text{otherwise} \end{cases} \tag{2.5}$$

Then for the edge $e = (s,t)$, we define its *slack* as:

$$\sigma_e(\mathbf{d}(\mathbf{x})) = R_t(\mathbf{d}(\mathbf{x})) - A_s(\mathbf{d}(\mathbf{x})) - d_e(\mathbf{x}) \tag{2.6}$$

For any net i, we define its slack as $\sigma_i(\mathbf{d}(\mathbf{x})) = \min_{e\in i} \sigma_e(\mathbf{d}(\mathbf{x}))$.

Now let $\mathcal{P}$, $\mathcal{P}_\mathcal{I}$, and $\mathcal{P}_\mathcal{O}$ denote the (finite) set of all pins, input pins, and output pins in the circuit, respectively. All paths $\pi = (\pi_1, \pi_2, \dots, \pi_p)$ begin with an input pin $\pi_1 \in \mathcal{P}_\mathcal{I}$, end with an output pin $\pi_p \in \mathcal{P}_\mathcal{O}$, and adjacent pins in the path are connected by an edge. We require that $\mathcal{P}_\mathcal{I} \cap \mathcal{P}_\mathcal{O} = \emptyset$, i.e. there are no single-pin paths. Let Π denote the set of all paths in the circuit, and let $S = |\Pi|$. Note that we can also think of each net i and path π as a set and string of edges, respectively, and do so when appropriate.

Now we are ready to present the problem to be solved. For simplicity, to prevent overlaps we temporarily assume some pins are fixed; the more practical case involving density constraints is fully compatible in this framework and discussed in section 3.2 below.

$$\begin{aligned}
&\min_{\mathbf{x}} \; G(\mathbf{x}) \\
&\textit{subject to:} \\
&\sigma_\pi(\mathbf{d}(\mathbf{x})) \geq 0 \quad \forall \text{ paths } \pi \in \Pi \\
&\text{(some pins fixed)}
\end{aligned} \tag{2.7}$$

Finally, one last piece of notation. Let Π_e denote the set of all paths containing edge e, and let $\Pi_i = \bigcup_{e \in i} \Pi_e$ denote the set of all paths passing through net i. Denote by $\mathfrak{F}$ the set of feasible placements $\mathbf{x}$ in the problem (2.7). Analogously, denote by $\mathfrak{F}_0$ the set of placements strictly feasible for all path timing constraints, i.e. the set of all $\mathbf{x}$ for which $\sigma_\pi(\mathbf{d}(\mathbf{x})) > 0$ for all paths $\pi \in \Pi$. We make the mild assumption that $\mathfrak{F}_0 \neq \emptyset$.

2.2 Net Weighting Framework

Now let us introduce the sequence of *weighted objective functions* $N^k(\mathbf{x})$ corresponding to the index $k = 1, 2, \dots$ as the following:

$$N^k(\mathbf{x}) = G(\mathbf{x}) + \Psi^k(\mathbf{x}) \tag{2.8}$$

where

$$\Psi^k(\mathbf{x}) = \sum_{\text{nets } i \in \mathcal{N}} w_i^k(\mathbf{d}(\mathbf{x}))h_i(\mathbf{x})$$

Here for each net i, $\{w_i^k\}_{k=1}^{\infty}$ is a sequence of *continuous, nonnegative* weighting functions which we require to satisfy the following property, termed the *asymptotic slack control*, for any fixed edge delay vector $\mathbf{d} = \mathbf{d}(\mathbf{x})$:

$$\lim_{k \to \infty} w_i^k(\mathbf{d}) = \begin{cases} 0, & \text{if } \sigma_i(\mathbf{d}) > 0. \\ c_i(\mathbf{d}), & \text{if } \sigma_i(\mathbf{d}) = 0, \text{ where } c_i(\mathbf{d}) \text{ is some finite constant.} \\ \infty, & \text{if } \sigma_i(\mathbf{d}) < 0. \end{cases}$$

(2.9) Asymptotic slack control.

Note that since $N^k(\mathbf{x}) = \sum_{\text{nets } i \in \mathcal{N}} [1 + w_i^k(\mathbf{x})]h_i(\mathbf{x})$, implementing the new objective amounts to simply multiplying each net measure $h_i(\mathbf{x})$ by the *net weight* $1 + w_i^k(\mathbf{x})$ in the original objective $G(\mathbf{x})$.

Please refer to Table A.1 in the Appendix for a full summary of notation used throughout this work.

2.3 Weighting Examples

We now identify particular weighting schemes that satisfy the sufficient criteria for convergence.

We can modify each of the VPR, PATH, and APlace weighting schemes, as defined in (1.1), (1.2), and (1.3), respectively, to satisfy the asymptotic slack control. To do this, we create *sequences* of weighting functions using an increasing *weighting parameter* α_k in place of the ad-hoc, user-defined parameter α. It should be noted that this change to satisfy the asymptotic slack control is necessary for convergence to the optimum: without such a modification, specific circuit examples can be found for which the original VPR, PATH, and APlace methods may fail to arrive at the optimum placement. Also, each of these three weighting schemes have been originally defined as *edge* weighting schemes; we may broaden their scope to *net* weighting schemes by summing over the edge weights, i.e. $w_i = \sum_{e \in i} w_e$.

We modify the VPR weights to the following, named *VPR-c*:

$$w_i^k(\mathbf{d}(\mathbf{x})) = \sum_{e \in i} \Big(1 - \frac{\sigma_e(\mathbf{d}(\mathbf{x}))}{T}\Big)^{\alpha_k} \tag{2.10}$$

Here $\{\alpha_k\}$ is a sequence of parameters approaching infinity, and the max delay of the previous placement, T_p, has been replaced with the desired max path delay T. This weighting scheme satisfies the asymptotic slack control.

We use the following modified formulation of the PATH weighting, which we name *PATH-c*:

$$w_i^k(\mathbf{d}(\mathbf{x})) = \sum_{\pi \in \Pi_i} \alpha_k^{-\frac{\sigma_\pi(\mathbf{d}(\mathbf{x}))}{T}} \tag{2.11}$$

Here, Π_i is the set of all paths passing through net i, and $\{\alpha_k\}$ is again a sequence of parameters approaching infinity, with $\alpha_k > 1$ for each k. This weighting scheme satisfies the asymptotic slack control. Figure 2.1 shows the shape of the PATH-c weighting function for increasing values of the net weighting parameter.

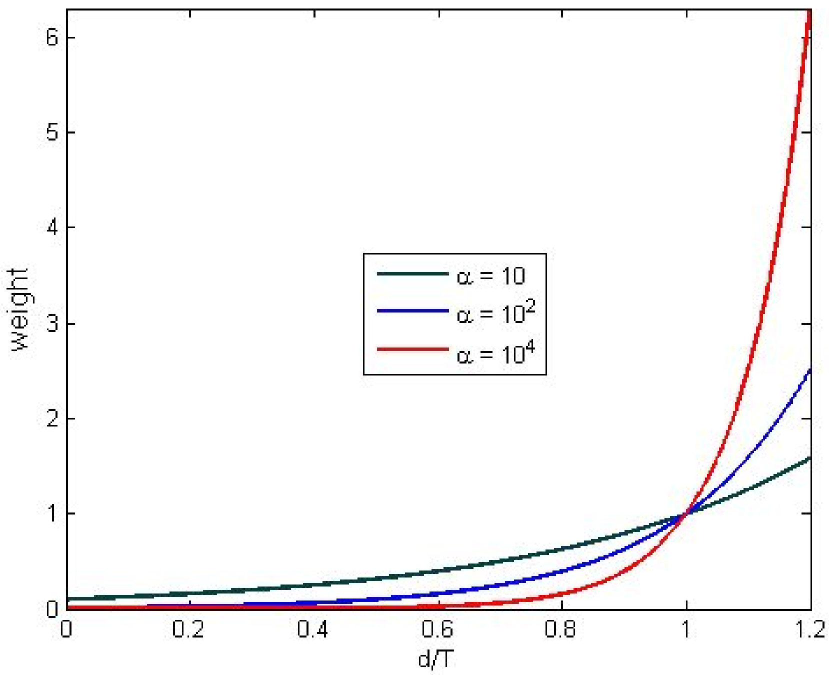

Figure 2.1: PATH-c weighting for increasing net weighting parameter α.

For the APlace weighting, we additionally replace the *desired timing improvement* T_u with the fixed desired max path delay T. The following modified weighting formulation is named *APlace-c*:

$$w_i^k(\mathbf{d}(\mathbf{x})) = \sum_{\pi \in \Pi_i} f^{\alpha_k}(d_\pi(\mathbf{x})),$$
$$\text{where } f^\alpha(d) = \begin{cases} \left(\frac{d}{T}\right)^\alpha - 1 & \text{if } d > T \\ 0 & \text{otherwise} \end{cases} \tag{2.12}$$

Here, $d_\pi(\mathbf{x}) = \sum_{e \in \pi} d_e(\mathbf{x})$, and $\{\alpha_k\}$ is a sequence of parameters approaching infinity. This weighting scheme satisfies the asymptotic slack control.

Note that the authors of the PATH and APlace weightings have devised efficient methods to calculate the weights, despite the exponential number paths through a net enumerated in their definitions. As described in [10], the PATH algorithm computes all weights in time linear to the number of pins plus the number of edges.

2.4 Preliminary Results

Before stating the main results of this paper, we present a necessary lemma and corollary to be used later.

Lemma 2.1. *For any edge delay vector* $\mathbf{d} = [d_1 \ \dots \ d_M]^T$, $\sigma_e(\mathbf{d}) = \min_{\pi \in \Pi_e}\{\sigma_\pi(\mathbf{d})\}$.

Proof. First we show that $\sigma_e(\mathbf{d}) \leq \sigma_\pi(\mathbf{d})$ for any $\pi \in \Pi_e$. Given $e = (s_1, t_1)$, take any $\pi = (s_p, \dots, s_2, s_1, t_1, t_2, \dots, t_q) \in \Pi_e$. Then by the definitions above we have

$$\begin{aligned}
\sigma_e(\mathbf{d}) &= R_{t_1}(\mathbf{d}) - A_{s_1}(\mathbf{d}) - d_{(s_1,t_1)} \\
&\leq R_{t_2}(\mathbf{d}) - d_{(t_1,t_2)} - A_{s_2}(\mathbf{d}) - d_{(s_2,s_1)} - d_{(s_1,t_1)} \\
&\vdots \\
&\leq R_{t_q}(\mathbf{d}) - A_{s_p}(\mathbf{d}) - \sum_{e_0 \in \pi} d_{e_0} \\
&= T - \sum_{e_0 \in \pi} d_{e_0} = \sigma_\pi(\mathbf{d})
\end{aligned} \tag{2.13}$$

Now, using a similar line of reasoning, given $e = (s_1, t_1)$ we show there exists a path $\pi^* \in \Pi_e$ such that $\sigma_e(\mathbf{d}) = \sigma_{\pi^*}(\mathbf{d})$. We do so by construction:

Start with $p = 1$, $q = 1$.
While $s_p \notin \mathcal{P}_\mathcal{I}$
  Choose $s_{p+1} \in \text{fanin}(s_p)$ such that $A_{s_{p+1}}(\mathbf{d}) + d_{(s_{p+1},s_p)} = A_{s_p}(\mathbf{d})$.
  Set $p \leftarrow p + 1$.
End while
While $t_q \notin \mathcal{P}_\mathcal{O}$
  Choose $t_{q+1} \in \text{fanout}(t_q)$ such that $R_{t_{q+1}}(\mathbf{d}) - d_{(t_q,t_{q+1})} = R_{t_q}(\mathbf{d})$.
  Set $q \leftarrow q + 1$.
End while
Set $\pi^* = (s_p, \dots, s_2, s_1, t_1, t_2, \dots, t_q)$.

It follows from construction of π^* that each inequality in (2.13) becomes an equality, so that $\sigma_e(\mathbf{d}) = \sigma_{\pi^*}(\mathbf{d})$. Combining this with the above statement that $\sigma_e(\mathbf{d}) \leq \sigma_\pi(\mathbf{d})$ for any $\pi \in \Pi_e$, we conclude that $\sigma_e(\mathbf{d}) = \min_{\pi \in \Pi_e}\{\sigma_\pi(\mathbf{d})\}$. □

Corollary 2.2. *The condition*

$$\text{If } \sigma_\pi(\mathbf{d}) > 0 \quad \forall \pi \in \Pi_i, \ \text{ then } \lim_{k\to\infty} w_i^k(\mathbf{d}) = 0. \tag{2.14a}$$

$$\text{If } \sigma_\pi(\mathbf{d}) \geq 0 \quad \forall \pi \in \Pi_i, \ \text{ then } \lim_{k\to\infty} w_i^k(\mathbf{d}) = c_i(\mathbf{d}), \text{ for some finite constant } c_i(\mathbf{d}). \tag{2.14b}$$

$$\text{If } \sigma_\pi(\mathbf{d}) < 0 \ \text{ for some } \pi \in \Pi_i, \ \text{ then } \lim_{k\to\infty} w_i^k(\mathbf{d}) = \infty. \tag{2.14c}$$

is equivalent to the asymptotic slack control (2.9).

This corollary will be useful in the following discussion.

3 Global Convergence

3.1 Proof of Convergence

Given this framework and net weightings defined by (2.8), we can guarantee convergence of a subsequence of the global minimizers of the weighted objectives $N^k(\mathbf{x})$ to a global minimizer of the original constrained problem.

Theorem 3.1. *Suppose that $\mathbf{x}^k$ is a global minimizer of the weighted objective $N^k(\mathbf{x})$ for each $k = 1, 2, \ldots$. Then every limit point of the sequence $\{\mathbf{x}^k\}$ is a global minimizer of the problem (2.7).*

Proof. Let $\mathbf{z}$ be a global minimizer of the problem (2.7). First note that since each $\mathbf{x}^k$ minimizes $N^k(\mathbf{x})$, we have that

$$N^k(\mathbf{x}^k) \leq N^k(\mathbf{z}), \text{ for all } k$$

i.e.,

$$G(\mathbf{x}^k) + \Psi^k(\mathbf{x}^k) \leq G(\mathbf{z}) + \Psi^k(\mathbf{z}), \text{ for all } k \tag{3.1}$$

Suppose $\mathbf{x}^*$ is a limit point of $\{\mathbf{x}^k\}$, so there exists some subsequence K such that $\lim_{k \in K} \mathbf{x}^k = \mathbf{x}^*$.

We first show that $\mathbf{x}^* \in \mathfrak{F}$. We have by continuity that $\sigma_\pi(\mathbf{d}(\mathbf{x}^*)) = \lim_{k \in K} \sigma_\pi(\mathbf{d}(\mathbf{x}^k))$ for each path π. Suppose that $\sigma_\pi(\mathbf{d}(\mathbf{x}^*)) < 0$ for some π. Then there there exists some Z such that for all $k \in K$ with $k > Z$,

$$\sigma_\pi(\mathbf{d}(\mathbf{x}^k)) < -\epsilon$$

for some $\epsilon > 0$. By property (2.14) it follows that for each net i through which π passes,

$$w_i^k(\mathbf{d}(\mathbf{x}^k)) \to \infty$$

Furthermore, since $\sigma_\pi(\mathbf{d}(\mathbf{x}^*)) < 0$ it follows by definition that $\sum_{e \in \pi} d_e(\mathbf{x}^*) > T > 0$, so there exists at least one $\hat{e} = (\hat{s}, \hat{t}) \in \pi$ such that $d_{\hat{e}}(\mathbf{x}^*) > 0$. Thus by (2.2), $x^*_{\hat{s}} \neq x^*_{\hat{t}}$ so that for the net $\hat{i} \ni \hat{e}$, it follows from (2.1) and nonnegativity that $h_{\hat{i}}(\mathbf{x}^*) > 0$. Then again by continuity we can choose $k \in K$ sufficiently large so that for some $R > 0$,

$$h_{\hat{i}}(\mathbf{x}^k) > R$$

Thus we get:

$$\begin{aligned}
\Psi^k(\mathbf{x}^k) &= \sum_{\text{nets } i \in \mathcal{N}} w_i^k(\mathbf{d}(\mathbf{x}^k)) h_i(\mathbf{x}^k) \\
&\geq w_{\hat{i}}^k(\mathbf{d}(\mathbf{x}^k)) h_{\hat{i}}(\mathbf{x}^k) \\
&\geq w_{\hat{i}}^k(\mathbf{d}(\mathbf{x}^k)) R \to \infty \text{ as } k \to \infty
\end{aligned} \tag{3.2}$$

Now, since $\mathbf{z} \in \mathfrak{F}$, it follows from (2.14) that $\lim_{k \in K} \Psi^k(\mathbf{z}) = C$ for some constant $C \geq 0$. This fact, combined with (3.2), contradicts (3.1) for sufficiently large $k \in K$. We thus conclude that $\sigma_\pi(\mathbf{d}(\mathbf{x}^*)) \geq 0$ for all paths π, so that $\mathbf{x}^* \in \mathfrak{F}$.

Now suppose that $\mathbf{x}^*$ is *not* a global minimizer of (2.7), i.e. $G(\mathbf{x}^*) > G(\mathbf{z})$. We will show that this implies that there exists a point $\mathbf{y} \in \mathfrak{F}_0$ such that $G(\mathbf{x}^*) > G(\mathbf{y})$.

If $\mathbf{z} \in \mathfrak{F}_0$, take $\mathbf{y} = \mathbf{z}$. Otherwise, choose any point $\tilde{\mathbf{x}} \in \mathfrak{F}_0$, and assume $G(\mathbf{x}^*) \leq G(\tilde{\mathbf{x}})$ (otherwise, take $\mathbf{y} = \tilde{\mathbf{x}}$). Now let $\mathbf{x}_\beta = \beta\tilde{\mathbf{x}} + (1-\beta)\mathbf{z}$ for $\beta \in (0,1)$. For each path π, we know that $\sigma_\pi(\mathbf{d}(\tilde{\mathbf{x}})) > 0$ and $\sigma_\pi(\mathbf{d}(\mathbf{z})) \geq 0$. It follows by convexity of the delay function that the pathwise slack function is concave; hence $\sigma_\pi(\mathbf{d}(\mathbf{x}_\beta)) \geq \beta\big(\sigma_\pi(\mathbf{d}(\tilde{\mathbf{x}}))\big) > 0$, so that $\mathbf{x}_\beta \in \mathfrak{F}_0$ for all $\beta \in (0,1)$.

Now it follows by continuity that we can choose $\beta = \beta_1$ sufficiently small so that $G(\mathbf{x}_{\beta_1}) < G(\mathbf{x}^*)$. We have found our desired point $\mathbf{y} = \mathbf{x}_{\beta_1}$.

Now recall that by definition of $\mathbf{x}^k$,

$$G(\mathbf{x}^k) + \Psi^k(\mathbf{x}^k) \leq G(\mathbf{y}) + \Psi^k(\mathbf{y}), \text{ for all } k$$

We can take the limit of both sides of this inequality to obtain

$$G(\mathbf{x}^*) + \lim_{k \in K} \Psi^k(\mathbf{x}^k) \leq G(\mathbf{y})$$

and so by nonnegativity of each w_i^k and h_i,

$$G(\mathbf{x}^*) \leq G(\mathbf{y}) \tag{3.3}$$

which contradicts our previous assertion that $G(\mathbf{x}^*) > G(\mathbf{y})$. We conclude that $\mathbf{x}^*$ is a global minimizer of (2.7). □

3.2 Extension to Generalized Density Constraints

Up to this point, we have made the simplifying assumption that some pins are fixed and there are no density or other additional constraints in the problem. We show in this section that the extension of problem (2.7) to include more general equality constraints does not interfere with the convergence of the net weighting scheme. The treatment of the density constraint as an equality constraint was conducted effectively in [4]. We choose to include a penalty term on the weighted objective and use standard techniques, as in e.g. [13], to show convergence.

Suppose we wish to solve the modified problem

$$\begin{aligned} &\min_{\mathbf{x}} \; G(\mathbf{x}) \\ &\textit{subject to:} \\ &\sigma_\pi(\mathbf{d}(\mathbf{x})) \geq 0 \quad \forall \text{ paths } \pi \in \Pi \\ &D_r(\mathbf{x}) = 0 \quad \forall r = 1, \ldots, R \end{aligned} \tag{3.4}$$

where the constraints $D_r(\mathbf{x}) = 0$ represent generalized density constraints, which could be used to account for overlap, routability, temperature, and so on. Traditionally, analytical placers (e.g., [15]) divide the placement area into a grid of "bins" and discourage overlapping cells by bounding the average density in each bin using an inequality constraint. In [4], filler "dummy" cells were introduced to convert the inequality constraints into equality constraints and were shown to be effective.

Let us redefine the weighted objective (2.8) to include additional penalty terms for the generalized density constraints:

$$\hat{N}^k(\mathbf{x}) = G(\mathbf{x}) + \Psi^k(\mathbf{x}) + \sum_{r=1}^{R} p_r^k\left(D_r(\mathbf{x})\right) \tag{3.5}$$

Here each p_r^k is a nonnegative function such that

$$\lim_{k \to \infty} p_r^k(D) = \begin{cases} 0 & \text{if } D = 0 \\ \infty & \text{otherwise} \end{cases}$$

Let $\mathfrak{D} = \{\mathbf{x} : D_r(\mathbf{x}) = 0,\ r = 1, \ldots, R\}$. We make the assumptions that $\mathfrak{F}_0 \cap \mathfrak{D} \neq \emptyset$, and that $\mathfrak{F} \cap \mathfrak{D}$ is in the closure of $\mathfrak{F}_0 \cap \mathfrak{D}$. It follows readily that we can modify Theorem 3.1 to include the penalty term:

Corollary 3.2. *Suppose that* $\mathbf{x}^k$ *is a global minimizer of the objective* $\hat{N}^k(\mathbf{x})$ *for each* $k = 1, 2, \ldots$. *Then every limit point of the sequence* $\{\mathbf{x}^k\}$ *is a global minimizer of the problem* (3.4).

Proof. We use the proof of Theorem 3.1 with a few modifications. First, note that the limit point $\mathbf{x}^* \in \mathfrak{D}$ in addition to $\mathbf{x}^* \in \mathfrak{F}$ as shown above, since otherwise $\sum_{r=1}^{R} p_r^k \left(D_r(\mathbf{x}^k)\right) \to \infty$, contradicting the fact that $\mathbf{x}^k$ minimizes $\hat{N}^k(\mathbf{x})$ for sufficiently large $k \in K$. Then, assuming that $\mathbf{x}^*$ is not a global minimum, we can find $\mathbf{y} \in \mathfrak{F}_0 \cap \mathfrak{D}$ such that $G(\mathbf{x}^*) > G(\mathbf{y})$, following from the fact that we can select points in $\mathfrak{F}_0 \cap \mathfrak{D}$ arbitrarily close to the global minimum $\mathbf{z}$, and use it to obtain a contradiction as in (3.3). □

4 Local Convergence

In many cases, given highly non-convex density constraints, finding global minimizers of the weighted objectives N^k is not practical. With this in mind, we now consider the case where we can only find a local minimizer for each subproblem. We show that doing so will still converge to a Karush-Kuhn-Tucker (KKT) point [13], i.e. a candidate point for a local minimizer, of an equivalent formulation of the original problem (2.7).

4.1 Preliminaries

4.1.1 Assumptions

We may retain all original assumptions about the functions d_e and h_i; however, the essential properties for local convergence are that for each edge e and net i, each h_i and d_e is *nonnegative*.

We also make the additional assumptions that the functions d_e and h_i are *continuously differentiable*, and that h_i is *Lipschitz continuous* in the feasible region $\mathfrak{F}$ and bounded below by some $\kappa > 0$. Note that Lipschitz continuity follows from continuous differentiability of h_i if $\mathfrak{F}$ is compact. Also, note that the lower bound κ on h_i holds for the log-sum-exp wire length definition; for the quadratic and half-perimeter wire lengths, it can be satisfied simply by making the adjustment $h_i^{\text{new}} = h_i + \kappa$.

4.1.2 Net Weighting Framework

As in Section 4.1.1, we may retain all original requirements on the net weighting functions w_i^k, but the essential properties for local convergence are that each w_i^k is *nonnegative*, and that the asymptotic slack control (2.9) is satisfied.

Further, we require that each w_i^k is *continuously differentiable* and make some additional requirements on their structure. First, we require that each w_i^k is *nondecreasing*:

$$\boxed{\nabla_{\mathbf{d}} w_i^k(\mathbf{d}(\mathbf{x})) \geq \mathbf{0}}$$

(4.1) Nondecreasing property.

Thus, increasing any edge delay while holding all other delays constant cannot have the effect of *reducing* a weight.

Secondly, we require that the weighting functions are *non-critically indifferent*: in the limit, changes in non-critical edge delays do not have an impact on the weights.

$$\text{If } \sigma_e(\mathbf{d}(\mathbf{x})) > 0, \text{ then } \lim_{k\to\infty} \frac{\partial w_i^k(\mathbf{d}(\mathbf{x}))}{\partial d_e} = 0 \quad \forall \text{ nets } i$$

(4.2) Non-critical indifference.

Also, we require that each w_i^k is *path consistent*: it can be written as some function of the path delays in the circuit.

$$w_i^k(\mathbf{d}(\mathbf{x})) = f_i^k\left(d_{\pi_1}(\mathbf{x}), \ldots, d_{\pi_S}(\mathbf{x})\right)$$

(4.3) Path consistency.

Finally, we make the mild expectation that the net weighting objectives $N^k(\mathbf{x})$ do not become arbitrarily "jagged" around local minimizers in $\mathfrak{F}$; that is, there exists a $\xi > 0$ independent of k such that if $\mathbf{y} \in \mathfrak{F}$ is a local minimizer for $N^k(\mathbf{x})$, then $N^k(\mathbf{x})$ is convex in $B_{\mathfrak{F}}(\mathbf{y}, \xi) = \{\mathbf{x} : \mathbf{x} \in \mathfrak{F},\ ||\mathbf{x} - \mathbf{y}|| < \xi\}$.

We can verify that each of the identified convergent weighting schemes VPR-c (2.10), PATH-c (2.11), and APlace-c (2.12) are indeed *nondecreasing*, *non-critically indifferent*, and *path consistent.*

The differentiability requirement is satisfied by the PATH-c weights, and may be satisfied by a slight modification in the VPR-c and APlace-c weights to smooth around non-differentiable points. The non-differentiability is due to the use of max and min functions in computing the slacks in the VPR-c weights, and due to the piecewise nature of the weighting function at the point $d = T$ in the APlace-c weights. The non-differentiable points in the VPR-c weightings can be removed if the log-sum-exp smooth approximations of the max and min functions [15] are used in computing the slacks. The non-differentiable points in the APlace-c weightings can be removed by a simple smoothing of f^a around the point $d = T$.

4.1.3 Problem Formulation

In an equivalent formulation to (2.7), we introduce the new variable $\mathbf{a} = (a_1, a_2, \ldots, a_n)$, where a_j is an upper bound on the arrival time of pin j. The problem then becomes:

$$\begin{aligned}
&\min_{\mathbf{x},\mathbf{a}}\ G(\mathbf{x}) \\
&\textit{subject to:} \\
&a_s + d_e(\mathbf{x}) \le a_t \quad \forall \text{ edges } e \in \mathcal{E}, \text{ where } e = (s,t) \\
&0 \le a_j \quad \forall \text{ pins } j \in \mathcal{P}_{\mathcal{I}} \\
&a_\ell \le T \quad \forall \text{ pins } \ell \in \mathcal{P}_{\mathcal{O}} \\
&\text{(some pins fixed)}
\end{aligned} \tag{4.4}$$

Here we denote the set of all edges by $\mathcal{E}$. Again we assume for simplicity that some pins are fixed; see Section 4.3 for inclusion of generalized density constraints. For this problem, the Lagrangian is:

$$\begin{aligned}
\mathcal{L}(\mathbf{x}, \mathbf{a}, \mathbf{u}, \mathbf{v}, \boldsymbol{\lambda}) = \sum_{\text{nets } i \in \mathcal{N}} h_i(\mathbf{x}) - \sum_{\text{pins } j \in \mathcal{P}_{\mathcal{I}}} u_j a_j - \sum_{\text{pins } \ell \in \mathcal{P}_{\mathcal{O}}} v_\ell (T - a_\ell) \\
- \sum_{\text{edges } e=(s,t)\in\mathcal{E}} \lambda_e (a_t - a_s - d_e(\mathbf{x}))
\end{aligned}$$

Then for any local minimum $(\bar{\mathbf{x}}, \bar{\mathbf{a}})$ for the problem (4.4) at which the linear independence constraint qualification [13] holds, by the KKT conditions there are Lagrange multipliers $(\bar{\mathbf{u}}, \bar{\mathbf{v}}, \bar{\boldsymbol{\lambda}})$ such that the following conditions are satisfied:

$$\sum_{e\in\mathcal{E}} \bar{\lambda}_e \nabla d_e(\bar{\mathbf{x}}) = -\sum_{i\in\mathcal{N}} \nabla h_i(\bar{\mathbf{x}}) \tag{4.5a}$$

$$\bar{u}_j = \sum_{e=(j,t)\in\mathcal{E}} \bar{\lambda}_e \qquad \forall j \in \mathcal{P}_\mathcal{I} \tag{4.5b}$$

$$\bar{a}_j \geq 0 \text{ and } \bar{u}_j \geq 0, \text{ with at least one of these a strict equality} \qquad \forall j \in \mathcal{P}_\mathcal{I} \tag{4.5c}$$

$$\bar{v}_\ell = \sum_{e=(s,\ell)\in\mathcal{E}} \bar{\lambda}_e \qquad \forall \ell \in \mathcal{P}_\mathcal{O} \tag{4.5d}$$

$$\bar{a}_\ell \leq T \qquad \forall \ell \in \mathcal{P}_\mathcal{O} \tag{4.5e}$$

$$\bar{v}_\ell \geq 0 \qquad \forall \ell \in \mathcal{P}_\mathcal{O} \tag{4.5f}$$

$$\bar{v}_\ell = 0 \;\text{ or }\; \bar{a}_\ell = T \qquad \forall \ell \in \mathcal{P}_\mathcal{O} \tag{4.5g}$$

$$\sum_{e_1=(s,m)\in\mathcal{E}} \bar{\lambda}_{e_1} = \sum_{e_2=(m,t)\in\mathcal{E}} \bar{\lambda}_{e_2} \qquad \forall m \in \mathcal{P} \setminus (\mathcal{P}_\mathcal{I} \cup \mathcal{P}_\mathcal{O}) \tag{4.5h}$$

$$\bar{a}_s + d_e(\mathbf{x}) \leq \bar{a}_t \qquad \forall e = (s,t) \in \mathcal{E} \tag{4.5i}$$

$$\bar{\lambda}_e \geq 0 \qquad \forall e \in \mathcal{E} \tag{4.5j}$$

$$\bar{\lambda}_e = 0 \;\text{ or }\; \bar{a}_s + d_e(\mathbf{x}) = \bar{a}_t \qquad \forall e = (s,t) \in \mathcal{E} \tag{4.5k}$$

4.2 Local Convergence Proof

Given a placement $\mathbf{x}$ and an index k, we shall define the variable $\tilde{\mathbf{a}}(\mathbf{d}(\mathbf{x}))$ and the *Lagrange multiplier estimates* $\tilde{\mathbf{u}}^k(\mathbf{x})$, $\tilde{\mathbf{v}}^k(\mathbf{x})$, and $\tilde{\boldsymbol{\lambda}}^k(\mathbf{x})$ as follows:

$$\tilde{a}_m(\mathbf{d}(\mathbf{x})) = A_m(\mathbf{d}(\mathbf{x})) \qquad \forall m \in \mathcal{P} \tag{4.6a}$$

$$\tilde{\lambda}_e^k(\mathbf{x}) = \sum_{i\in\mathcal{N}} \tfrac{\partial w_i^k(\mathbf{d}(\mathbf{x}))}{\partial d_e} h_i(\mathbf{x}) \qquad \forall e \in \mathcal{E} \tag{4.6b}$$

$$\tilde{u}_j^k(\mathbf{x}) = \sum_{e=(j,t)\in\mathcal{E}} \tilde{\lambda}_e^k(\mathbf{x}) \qquad \forall j \in \mathcal{P}_\mathcal{I} \tag{4.6c}$$

$$\tilde{v}_\ell^k(\mathbf{x}) = \sum_{e=(s,\ell)\in\mathcal{E}} \tilde{\lambda}_e^k(\mathbf{x}) \qquad \forall \ell \in \mathcal{P}_\mathcal{O} \tag{4.6d}$$

Theorem 4.1. *Suppose that $\mathbf{x}^k$ is a local minimizer of the weighted objective $N^k(\mathbf{x})$ for each $k = 1, 2, \ldots,$ and that $\mathbf{x}^k \in \mathfrak{F}_0$ for all k sufficiently large. Then every limit point of $\{\mathbf{x}^k\}$ is a KKT point of the problem* (4.4).

Proof. As in the proof of Theorem 3.1, let $\mathbf{x}^*$ be a limit point of $\{\mathbf{x}^k\}$, so there exists some subsequence K such that $\lim_{k\in K} \mathbf{x}^k = \mathbf{x}^*$. The proof will proceed by considering the variable $\tilde{\mathbf{a}}(\mathbf{d}(\mathbf{x}^k))$ and the estimates $\tilde{\boldsymbol{\lambda}}^k(\mathbf{x}^k)$, $\tilde{\mathbf{u}}^k(\mathbf{x}^k)$, and $\tilde{\mathbf{v}}^k(\mathbf{x}^k)$ given by (4.6) and showing that these satisfy the KKT conditions (4.5) in the limit.

Properties (4.5b), (4.5d), and (4.5i) are clearly satisfied at every iteration via construction. Property (4.5j) is also satisfied at every iteration, following from the nondecreasing property (4.1). Property (4.5c) is satisfied at every iteration as well: this follows from the aforementioned fact that $\tilde{\lambda}_e^k(\mathbf{x}) \geq 0$ for all k and $e \in \mathcal{E}$, and from the fact that $\tilde{a}_m(\mathbf{d}(\mathbf{x})) = A_m(\mathbf{d}(\mathbf{x})) = 0$ for all $m \in \mathcal{P}_\mathcal{I}$. Similarly, Property (4.5f) is satisfied at every iteration.

For Property (4.5h), first note that for any $m \in \mathcal{P} \setminus (\mathcal{P}_\mathcal{I} \cup \mathcal{P}_\mathcal{O})$,

$$\bigcup_{e_1=(s,m)\in\mathcal{E}} \Pi_{e_1} = \bigcup_{e_2=(m,t)\in\mathcal{E}} \Pi_{e_2}$$

Also, from path consistency (4.3) we have

$$\frac{\partial w_i^k(\mathbf{d})}{\partial d_e} = \sum_{\pi\in\Pi_e} \frac{\partial f_i^k\big(d_{\pi_1}, \ldots, d_{\pi_S}\big)}{\partial d_\pi}$$

Combining these, we get

$$\begin{aligned}
\sum_{e_1=(s,m)\in\mathcal{E}} \tilde{\lambda}_{e_1}^k(\mathbf{x}) &= \sum_{e_1=(s,m)\in\mathcal{E}} \left[\sum_{i\in\mathcal{N}} \frac{\partial w_i^k(\mathbf{d}(\mathbf{x}))}{\partial d_{e_1}} h_i(\mathbf{x})\right] \\
&= \sum_{e_1=(s,m)\in\mathcal{E}} \left[\sum_{i\in\mathcal{N}} \left(\sum_{\pi\in\Pi_{e_1}} \frac{\partial f_i^k\big(d_{\pi_1}(\mathbf{x}), \ldots, d_{\pi_S}(\mathbf{x})\big)}{\partial d_\pi} h_i(\mathbf{x})\right)\right] \\
&= \sum_{e_2=(m,t)\in\mathcal{E}} \left[\sum_{i\in\mathcal{N}} \left(\sum_{\pi\in\Pi_{e_2}} \frac{\partial f_i^k\big(d_{\pi_1}(\mathbf{x}), \ldots, d_{\pi_S}(\mathbf{x})\big)}{\partial d_\pi} h_i(\mathbf{x})\right)\right] \\
&= \sum_{e_2=(m,t)\in\mathcal{E}} \tilde{\lambda}_{e_2}^k(\mathbf{x})
\end{aligned}$$

Thus Property (4.5h) is satisfied at every iteration.

The complementary slackness condition (4.5k) is satisfied in the limit: since each $\mathbf{x}^k$ is timing feasible for k sufficiently large, it follows that $A_t(\mathbf{d}(\mathbf{x}^k)) \leq R_t(\mathbf{d}(\mathbf{x}^k))$ for all k sufficiently large and all pins $t \in \mathcal{P}$. Consider any edge $e = (s,t) \in \mathcal{E}$ such that $\tilde{a}_s(\mathbf{d}(\mathbf{x}^*)) + d_e(\mathbf{x}^*) < \tilde{a}_t(\mathbf{d}(\mathbf{x}^*))$; thus, $\tilde{a}_s(\mathbf{d}(\mathbf{x}^k)) + d_e(\mathbf{x}^k) < \tilde{a}_t(\mathbf{d}(\mathbf{x}^k)) - \epsilon$ for some $\epsilon > 0$ for all k sufficiently large. Then

$$\sigma_e(\mathbf{d}(\mathbf{x}^k)) = R_t(\mathbf{d}(\mathbf{x}^k)) - A_s(\mathbf{d}(\mathbf{x}^k)) - d_e(\mathbf{x}^k) \geq A_t(\mathbf{d}(\mathbf{x}^k)) - A_s(\mathbf{d}(\mathbf{x}^k)) - d_e(\mathbf{x}^k) > \epsilon$$

Thus $\sigma_e(\mathbf{d}(\mathbf{x}^*)) > \epsilon$ so it follows from non-critical indifference (4.2) and continuity that $\lim_{k\in K} \tilde{\lambda}_e^k(\mathbf{x}^k) = 0$.

Properties (4.5e) and (4.5g) are shown in a similar manner to Property (4.5k). It follows from timing feasibility for k sufficiently large that Property (4.5e) is satisfied in the limit. For Property (4.5g), consider any $\ell \in \mathcal{P}_\mathcal{O}$ such that $\tilde{a}_\ell(\mathbf{d}(\mathbf{x}^*)) < T$, so that $\tilde{a}_\ell(\mathbf{d}(\mathbf{x}^k)) < T - \epsilon$ for some $\epsilon > 0$ for all k sufficiently large. Now note that for any edge $(s,\ell) \in \mathcal{E}$, we have that $A_\ell(\mathbf{d}(\mathbf{x}^k)) \geq A_s(\mathbf{d}(\mathbf{x}^k)) + d_{(s,\ell)}(\mathbf{x}^k)$, and since $\ell \in \mathcal{P}_\mathcal{O}$, we have $R_\ell(\mathbf{d}(\mathbf{x}^k)) = T$. Thus,

$$\sigma_{(s,\ell)}(\mathbf{d}(\mathbf{x}^k)) = T - A_s(\mathbf{d}(\mathbf{x}^k)) - d_{(s,\ell)}(\mathbf{x}^k) \geq T - A_\ell(\mathbf{d}(\mathbf{x}^k)) > \epsilon$$

Thus again by non-critical indifference (4.2), it follows that $\lim_{k\in K} \tilde{v}_\ell^k(\mathbf{x}^k) = 0$.

Finally we prove Property (4.5a) in the limit. First we show that $w_i^k(\mathbf{d}(\mathbf{x}^k)) \to 0$ for all $i \in \mathcal{N}$. As a means to show this, denote by $\mathcal{N}_+$ the set of all nets $i \in \mathcal{N}$ such that there exists some $\epsilon > 0$ such that $w_i^k(\mathbf{d}(\mathbf{x}^k)) > \epsilon$ for all $k \in K$ sufficiently large. Further, let $\mathcal{N}_0 = \mathcal{N} \setminus \mathcal{N}_+$.

Suppose temporarily that $\mathcal{N}_+ \neq \emptyset$. Then given an arbitrarily small $0 < \delta_1 < \epsilon$ and letting $\delta = \min\{\delta_1, \xi\}$, we can find an index q and a point $\mathbf{y} \in \mathfrak{F}_0$ such that

$$\begin{aligned}
&||\mathbf{y} - \mathbf{x}^q|| < \delta \\
&w_i^q(\mathbf{d}(\mathbf{y})) < \delta \quad \forall i \in \mathcal{N}
\end{aligned}$$

q and $\mathbf{y}$ can be found by first selecting an M sufficiently large and $\mathbf{y} \in \mathfrak{F}_0$ sufficiently close to $\mathbf{x}^*$ so that $\mathbf{x}^k \in \mathfrak{F}_0$ and $\big|\big|\mathbf{x}^k - \mathbf{y}\big|\big| < \delta$ for all $k > M$, then fixing $\mathbf{y}$ and increasing $q > M$ until $w_i^q(\mathbf{d}(\mathbf{y})) < \delta$

for all $i \in \mathcal{N}$. Then we have

$$\begin{aligned} N^q(\mathbf{x}^q) &> \sum_{i\in\mathcal{N}} h_i(\mathbf{x}^q) + \epsilon \sum_{j\in\mathcal{N}_+} h_j(\mathbf{x}^q) \\ &= \sum_{i\in\mathcal{N}} h_i(\mathbf{x}^q) + \delta \sum_{j\in\mathcal{N}_+} h_j(\mathbf{x}^q) + (\epsilon - \delta) \sum_{j\in\mathcal{N}_+} h_j(\mathbf{x}^q) \\ &\geq \sum_{i\in\mathcal{N}} h_i(\mathbf{x}^q) + \delta \sum_{j\in\mathcal{N}_+} h_j(\mathbf{x}^q) + (\epsilon - \delta) \left|\mathcal{N}_+\right| \kappa \end{aligned} \tag{4.7}$$

Now by Lipschitz continuity, there exists some constant L such that $|h_i(\mathbf{x}) - h_i(\mathbf{z})| \leq L \left|\left|\mathbf{x} - \mathbf{z}\right|\right|$ for all $i \in \mathcal{N}$ and $\mathbf{x}, \mathbf{z} \in \mathfrak{F}$. Thus $|h_i(\mathbf{y}) - h_i(\mathbf{x}^q)| \leq L\delta$, so

$$h_i(\mathbf{y}) \leq h_i(\mathbf{x}^q) + L\delta$$

for all $i \in \mathcal{N}$. Then we have:

$$\begin{aligned} N^q(\mathbf{y}) &< \sum_{i\in\mathcal{N}} \left[h_i(\mathbf{x}^q) + L\delta\right] + \delta \sum_{j\in\mathcal{N}_+} \left[h_j(\mathbf{x}^q) + L\delta\right] + \delta \sum_{j\in\mathcal{N}_0} H \\ &= \sum_{i\in\mathcal{N}} h_i(\mathbf{x}^q) + \delta \sum_{j\in\mathcal{N}_+} h_j(\mathbf{x}^q) + \delta \left|\mathcal{N}\right| L + \delta^2 \left|\mathcal{N}_+\right| L + \delta \left|\mathcal{N}_0\right| H \end{aligned} \tag{4.8}$$

where $H = \max\{h_i(\mathbf{x}) : i \in \mathcal{N}, \mathbf{x} \in \mathfrak{F}\}$ is a positive real number, independent of q or $\mathbf{y}$. Thus it follows from (4.7) and (4.8) that we can choose δ_1 sufficiently small to obtain q and $\mathbf{y}$ such that $N^q(\mathbf{y}) < N^q(\mathbf{x}^q)$ for $||\mathbf{y} - \mathbf{x}^q|| < \xi$, a contradiction of the fact that $\mathbf{x}^q$ is a local minimizer of $N^q(\mathbf{x})$. We conclude that $\mathcal{N}_+ = \emptyset$ so that $w_i^k(\mathbf{d}(\mathbf{x}^k)) \to 0$ for all $i \in \mathcal{N}$.

Next, let us restate our definition of the weighted objective $N^k(\mathbf{x})$ in a slightly modified form:

$$N^k(\mathbf{x}) = \sum_{\text{nets } i\in\mathcal{N}} \Big(1 + w_i^k(\mathbf{d}(\mathbf{x}))\Big) h_i(\mathbf{x})$$

Taking the gradient of this, we get

$$\begin{aligned} \nabla N^k(\mathbf{x}) &= \sum_{i\in\mathcal{N}} \Big(1 + w_i^k(\mathbf{d}(\mathbf{x}))\Big) \nabla h_i(\mathbf{x}) + \sum_{i\in\mathcal{N}} \left[\sum_{e\in\mathcal{E}} \tfrac{\partial w_i^k(\mathbf{d}(\mathbf{x}))}{\partial d_e} \nabla d_e(\mathbf{x})\right] h_i(\mathbf{x}) \\ &= \sum_{i\in\mathcal{N}} \Big(1 + w_i^k(\mathbf{d}(\mathbf{x}))\Big) \nabla h_i(\mathbf{x}) + \sum_{e\in\mathcal{E}} \left[\sum_{i\in\mathcal{N}} \tfrac{\partial w_i^k(\mathbf{d}(\mathbf{x}))}{\partial d_e} h_i(\mathbf{x})\right] \nabla d_e(\mathbf{x}) \\ &= \sum_{i\in\mathcal{N}} \Big(1 + w_i^k(\mathbf{d}(\mathbf{x}))\Big) \nabla h_i(\mathbf{x}) + \sum_{e\in\mathcal{E}} \tilde{\lambda}_e^k(\mathbf{x}) \nabla d_e(\mathbf{x}) \end{aligned} \tag{4.9}$$

Now, since $w_i^k(\mathbf{d}(\mathbf{x}^k)) \to 0$ for all $i \in \mathcal{N}$ and since $\nabla N^k(\mathbf{x}) = 0$, it follows from (4.9) that Property (4.5a) is satisfied in the limit. As all KKT conditions are satisfied, we conclude that $\mathbf{x}^*$ is a KKT point of (4.4). □

The following corollary shows that we need not find local minimizers of the weighted objectives $N^k(\mathbf{x})$ *exactly*, but may instead use approximations.

Corollary 4.2. *Suppose that $\mathbf{x}^k$ approximates a local minimizer $\mathbf{x}^k_{min}$ of the weighted objective $N^k(\mathbf{x})$ for each $k = 1, 2, \ldots$, in that $\left|\left|\mathbf{x}^k - \mathbf{x}^k_{min}\right|\right| \to 0$ and $\left|\left|\nabla N^k(\mathbf{x}^k)\right|\right| \to 0$ as $k \to \infty$. If $\mathbf{x}^k \in \mathfrak{F}_0$ and $\mathbf{x}^k_{min} \in \mathfrak{F}$ for all k sufficiently large, then every limit point of $\{\mathbf{x}^k\}$ is a KKT point of the problem* (4.4).

Proof. The proof proceeds as that for Theorem 4.1 with a few modifications. We discuss Property (4.5a) fully.

First we show that $w_i^k(\mathbf{d}(\mathbf{x}^k)) \to 0$ for all $i \in \mathcal{N}$. Suppose temporarily that $\mathcal{N}_+ \neq \emptyset$. Then given an arbitrarily small $0 < \delta_1 < \epsilon$ and letting $\delta = \min\{\delta_1, \frac{\xi}{2}\}$, we can find an index q and a point $\mathbf{y} \in \mathfrak{F}_0$ such that

$$\begin{aligned} &||\mathbf{y} - \mathbf{x}^q|| < \delta \\ &||\nabla N^q(\mathbf{x}^q)|| < 1 \\ &||\mathbf{x}^q - \mathbf{x}^q_{\min}|| < \delta \\ &w_i^q(\mathbf{d}(\mathbf{y})) < \delta \quad \forall i \in \mathcal{N} \end{aligned}$$

$\mathbf{y}$ and q can be found by first selecting an M sufficiently large and $\mathbf{y}$ sufficiently close to $\mathbf{x}^*$ so that $\mathbf{x}^k \in \mathfrak{F}_0$, $\mathbf{x}^k_{\min} \in \mathfrak{F}$, $\left|\left|\mathbf{y} - \mathbf{x}^k\right|\right| < \delta$, $\left|\left|\nabla N^k(\mathbf{x}^k)\right|\right| < 1$, and $\left|\left|\mathbf{x}^k - \mathbf{x}^k_{\min}\right|\right| < \delta$ for all $f \in K$ such that $k > M$, then fixing $\mathbf{y}$ and increasing $q > M$ until $w_i^q(\mathbf{d}(\mathbf{y})) < \delta$ for all $i \in \mathcal{N}$.

Now note that it follows from Taylor's theorem and the Cauchy-Schwarz inequality that

$$|N^q(\mathbf{x}^q) - N^q(\mathbf{x}^q_{\min})| = \left|(\mathbf{x}^q - \mathbf{x}^q_{\min})^T \nabla N^q(\boldsymbol{\chi}^q)\right| \leq ||\mathbf{x}^q - \mathbf{x}^q_{\min}||\, ||\nabla N^q(\boldsymbol{\chi}^q)|| \tag{4.10}$$

for some $\boldsymbol{\chi}^q = \psi \mathbf{x}^q + (1-\psi)\mathbf{x}^q_{\min}$, where $\psi \in (0,1)$. Then we have that $\boldsymbol{\chi}^q \in \mathfrak{F}_0$ by convexity of the delay function. Thus since $\boldsymbol{\chi}^q \in B_{\mathfrak{F}}(\mathbf{x}^q_{\min}, \xi)$, it follows from non-jaggedness and from $||\nabla N^q(\mathbf{x}^q)|| < 1$ and $||\nabla N^q(\mathbf{x}^q_{\min})|| = 0$ that $||\nabla N^q(\boldsymbol{\chi}^q)|| < 1$. Thus, using (4.10) and $||\mathbf{x}^q - \mathbf{x}^q_{\min}|| < \delta$, we get

$$|N^q(\mathbf{x}^q) - N^q(\mathbf{x}^q_{\min})| < \delta$$

so that, in particular, $N^q(\mathbf{x}^q_{\min}) > N^q(\mathbf{x}^q) - \delta$.

Then we have

$$\begin{aligned} N^q(\mathbf{x}^q_{\min}) &> N^q(\mathbf{x}^q) - \delta \\ &> \sum_{i\in\mathcal{N}} h_i(\mathbf{x}^q) + \epsilon \sum_{j\in\mathcal{N}_+} h_j(\mathbf{x}^q) - \delta \\ &= \sum_{i\in\mathcal{N}} h_i(\mathbf{x}^q) + \delta \sum_{j\in\mathcal{N}_+} h_j(\mathbf{x}^q) + (\epsilon - \delta) \sum_{j\in\mathcal{N}_+} h_j(\mathbf{x}^q) - \delta \\ &\geq \sum_{i\in\mathcal{N}} h_i(\mathbf{x}^q) + \delta \sum_{j\in\mathcal{N}_+} h_j(\mathbf{x}^q) + (\epsilon - \delta)\, |\mathcal{N}_+|\, \kappa - \delta \end{aligned} \tag{4.11}$$

Now by Lipschitz continuity, there exists some constant L such that $|h_i(\mathbf{x}) - h_i(\mathbf{z})| \leq L\, ||\mathbf{x} - \mathbf{z}||$ for all $i \in \mathcal{N}$ and $\mathbf{x}, \mathbf{z} \in \mathfrak{F}$. Thus $|h_i(\mathbf{y}) - h_i(\mathbf{x}^q)| \leq L\delta$, so

$$h_i(\mathbf{y}) \leq h_i(\mathbf{x}^q) + L\delta$$

for all $i \in \mathcal{N}$. Then we have:

$$\begin{aligned} N^q(\mathbf{y}) &< \sum_{i\in\mathcal{N}} \left[h_i(\mathbf{x}^q) + L\delta\right] + \delta \sum_{j\in\mathcal{N}_+} \left[h_j(\mathbf{x}^q) + L\delta\right] + \delta \sum_{j\in\mathcal{N}_0} H \\ &= \sum_{i\in\mathcal{N}} h_i(\mathbf{x}^q) + \delta \sum_{j\in\mathcal{N}_+} h_j(\mathbf{x}^q) + \delta\, |\mathcal{N}|\, L + \delta^2\, |\mathcal{N}_+|\, L + \delta\, |\mathcal{N}_0|\, H \end{aligned} \tag{4.12}$$

where $H = \max\{h_i(\mathbf{x}) : i \in \mathcal{N}, \mathbf{x} \in \mathfrak{F}\}$ is a positive real number, independent of $\mathbf{y}$ and q. Thus it follows from (4.11) and (4.12) that we can choose δ_1 sufficiently small to obtain $\mathbf{y}$ and q such that $N^q(\mathbf{y}) < N^q(\mathbf{x}^q_{\min})$ for $||\mathbf{y} - \mathbf{x}^q_{\min}|| \leq ||\mathbf{y} - \mathbf{x}^q|| + ||\mathbf{x}^q - \mathbf{x}^q_{\min}|| < \xi$, a contradiction of the fact that $\mathbf{x}^q_{\min}$ is a local minimizer of $N^q(\mathbf{x})$. We conclude that $\mathcal{N}_+ = \emptyset$ so that $w_i^k(\mathbf{d}(\mathbf{x}^k)) \to 0$ for all $i \in \mathcal{N}$.

Next, taking the gradient of the weighted objective $N^k(\mathbf{x})$ exactly as above in (4.9), we obtain

$$\nabla N^k(\mathbf{x}) = \sum_{i\in\mathcal{N}} \Big(1 + w_i^k(\mathbf{d}(\mathbf{x}))\Big) \nabla h_i(\mathbf{x}) + \sum_{e\in\mathcal{E}} \tilde{\lambda}^k_e(\mathbf{x}) \nabla d_e(\mathbf{x}) \tag{4.13}$$

Now, since $w_i^k(\mathbf{d}(\mathbf{x}^k)) \to 0$ for all $i \in \mathcal{N}$ and since $\nabla N^k(\mathbf{x}) \to 0$, it follows from (4.13) that Property (4.5a) is satisfied in the limit.

□

4.3 Extension to Generalized Density Equality Constraints

As in the global convergence proof, we will show extensibility of Theorem 4.1 to generalized density equality constraints. We find that in handling these constraints by a standard penalty function, the framework does not change in a fundamental manner.

With the introduction of generalized density equality constraints the problem (4.4) becomes:

$$\begin{aligned} &\min_{\mathbf{x},\mathbf{a}} \; G(\mathbf{x}) \\ &\textit{subject to:} \\ &a_s + d_e(\mathbf{x}) \le a_t \quad \forall \text{ edges } e \in \mathcal{E}, \text{ where } e = (s,t) \\ &0 \le a_j \quad \forall \text{ pins } j \in \mathcal{P}_\mathcal{I} \\ &a_\ell \le T \quad \forall \text{ pins } \ell \in \mathcal{P}_\mathcal{O} \\ &D_r(\mathbf{x}) = 0 \quad \forall r = 1, \ldots, R \end{aligned} \tag{4.14}$$

We introduce the new Lagrange multiplier $\boldsymbol{\tau}$, and the Lagrangian becomes

$$\begin{aligned} \breve{\mathcal{L}}(\mathbf{x},\mathbf{a},\mathbf{u},\mathbf{v},\boldsymbol{\lambda},\boldsymbol{\tau}) = &\sum_{\text{nets } i \in \mathcal{N}} h_i(\mathbf{x}) - \sum_{\text{pins } j \in \mathcal{P}_\mathcal{I}} u_j a_j - \sum_{\text{pins } \ell \in \mathcal{P}_\mathcal{O}} v_\ell (T - a_\ell) \\ &- \sum_{\text{edges } e=(s,t)\in\mathcal{E}} \lambda_e (a_t - a_s - d_e(\mathbf{x})) - \sum_{r=1}^{R} \tau_r D_r(\mathbf{x}) \end{aligned}$$

The KKT conditions (4.5) remain unchanged, with the exception that condition (4.5a) becomes

$$\sum_{e\in\mathcal{E}} \bar{\lambda}_e \nabla d_e(\bar{\mathbf{x}}) = -\sum_{i\in\mathcal{N}} \nabla h_i(\bar{\mathbf{x}}) + \sum_{r=1}^{R} \bar{\tau}_r \nabla D_r(\bar{\mathbf{x}}) \tag{4.15}$$

and we have the additional condition of feasibility with respect to the density constraints:

$$D_r(\mathbf{x}) = 0, \qquad r = 1, \ldots, R \tag{4.16}$$

We introduce a modified form of the penalty function as compared to (3.5):

$$\breve{N}^k(\mathbf{x}) = G(\mathbf{x}) + \Psi^k(\mathbf{x}) + \sum_{r=1}^{R} \mu^k \breve{p}_r \left(D_r(\mathbf{x})\right)$$

Here the penalty terms have the form $\breve{p}_r^k = \mu^k \breve{p}_r$, where μ^k is a sequence of positive real numbers such that $\mu^k \nearrow \infty$. The functions $\breve{p}_r$ are *nonnegative, continuously differentiable*, and have the properties

$$\breve{p}_r(D) = 0 \iff D = 0 \tag{4.17}$$

$$(\breve{p}_r)'(D) = 0 \implies D = 0 \tag{4.18}$$

Corollary 4.3. *Suppose that $\mathbf{x}^k$ is a local minimizer of the function $\breve{N}^k(\mathbf{x})$ for each $k = 1, 2, \ldots$, and that $\mathbf{x}^k \in \mathfrak{F}_0$ for all k sufficiently large. Then every limit point of the sequence $\{\mathbf{x}^k\}$ at which the linear independence constraint qualification (LICQ) [14] holds is a KKT point of the problem* (4.14)*.*

Proof. First we show Property (4.16). Corresponding to (4.9) we have

$$\nabla \breve{N}^k(\mathbf{x}) = \sum_{i\in\mathcal{N}} \Big(1 + w_i^k(\mathbf{d}(\mathbf{x}))\Big) \nabla h_i(\mathbf{x}) + \sum_{e\in\mathcal{E}} \tilde{\lambda}_e^k(\mathbf{x}) \nabla d_e(\mathbf{x}) + \mu^k \sum_{r=1}^{R} (\breve{p}_r)' \left(D_r(\mathbf{x})\right) \nabla D_r(\mathbf{x})$$

Since each $\mathbf{x}^k$ is a local minimizer of $\breve{N}^k(\mathbf{x})$, we have that $\nabla \breve{N}^k(\mathbf{x}^k) = 0$. Thus for each $k \in K$,

$$\sum_{r=1}^{R} (\breve{p}_r)' \left(D_r(\mathbf{x}^k)\right) \nabla D_r(\mathbf{x}^k) = \frac{1}{\mu^k} \left[-\sum_{i \in \mathcal{N}} \Big(1 + w_i^k(\mathbf{d}(\mathbf{x}^k))\Big) \nabla h_i(\mathbf{x}^k) - \sum_{e \in \mathcal{E}} \tilde{\lambda}_e^k(\mathbf{x}^k) \nabla d_e(\mathbf{x}^k) \right]$$

Taking the limit of both sides, we get

$$\sum_{r=1}^{R} (\breve{p}_r)' \left(D_r(\mathbf{x}^*)\right) \nabla D_r(\mathbf{x}^*) = 0$$

so that, provided the LICQ holds at $\mathbf{x}^*$, we have that $(\breve{p}_r)' \left(D_r(\mathbf{x}^*)\right) = 0$ for all $r = 1, \ldots, R$. Thus it follows from (4.18) that $D_r(\mathbf{x}^*) = 0$ for all $r = 1, \ldots, R$, so that Property (4.16) is satisfied in the limit.

To show Property (4.15), we introduce the Lagrange multiplier estimate $\tilde{\boldsymbol{\tau}}^k(\mathbf{x})$, in addition to those in (4.6):

$$\tilde{\tau}_j^k(\mathbf{x}) = -\mu^k (\breve{p}_r)' \left(D_r(\mathbf{x})\right) \qquad \forall r = 1, \ldots, R \tag{4.19}$$

It readily follows from the proof of Theorem 4.1 and the fact that $\nabla \breve{N}^k(\mathbf{x}^k) = 0$ for every k that Property (4.15) is satisfied. Note in particular that we may choose $\mathbf{y} \in \mathfrak{D}$ in addition to the properties outlined in the proof of Theorem 4.1 as a consequence of the fact that $\mathfrak{F} \cap \mathfrak{D}$ is in the closure of $\mathfrak{F}_0 \cap \mathfrak{D}$.

The remaining KKT conditions can be shown to be satisfied as a straightforward extension of the proof of Theorem 4.1. We conclude that $\mathbf{x}^*$ is a KKT point of (4.14). □

We may also extend Corollary 4.2 to account for the generalized density constraints.

Corollary 4.4. *Suppose that $\mathbf{x}^k$ approximates a local minimizer $\mathbf{x}^k_{min}$ of the weighted objective $\breve{N}^k(\mathbf{x})$ for each $k = 1, 2, \ldots,$ in that $\left|\left|\mathbf{x}^k - \mathbf{x}^k_{min}\right|\right| \to 0$ and $\left|\left|\nabla N^k(\mathbf{x}^k)\right|\right| \to 0$ as $k \to \infty$. If $\mathbf{x}^k \in \mathfrak{F}_0$ and $\mathbf{x}^k_{min} \in \mathfrak{F}$ for all k sufficiently large, then every limit point of $\{\mathbf{x}^k\}$ at which the LICQ holds is a KKT point of the problem* (4.14)*.*

Proof. The extension of Corollary 4.2 to the above is nearly identical to that in the proof of Corollary 4.3. We choose the new Lagrange multiplier estimate (4.19) as above and use it to show that Property (4.15) is satisfied. Similarly, we may use the same limit argument as above using the LICQ to show that Property (4.16) is satisfied in the limit. It is a straightforward extension of the proofs of Theorem 4.1 and Corollary 4.2 to show that the remaining KKT conditions are satisfied. □

4.4 Non-instantaneous Weighting Updates

We present one final result that relates to the practical implementation of this net weighting framework. Until now, we have made the assumption that the net weights are *continuously* updated according to the current placement $\mathbf{x}$. Normally in practice, we rely on a previous placement $\mathbf{y}$ to compute a set of net weights, then minimize using those fixed weights. Symbolically, we minimize the modified objective function

$$\breve{N}^k_{\mathbf{d}(\mathbf{y})}(\mathbf{x}) = \sum_{i \in \mathcal{N}} \Big(1 + w_i^k(\mathbf{d}(\mathbf{y}))\Big) h_i(\mathbf{x}) + \sum_{r=1}^{R} \mu^k \breve{p}_r \left(D_r(\mathbf{x})\right)$$

over $\mathbf{x}$. Using the previous iterate $\mathbf{y} = \mathbf{x}^{k-1}$ to set the net weights, we get the following result. Note that this is a revised version of Theorem 4.2 in [5].

Corollary 4.5. *Given any $\mathbf{x}^0$, suppose that $\mathbf{x}^k$ is a local minimizer of the weighted objective $\breve{N}^k_{\mathbf{d}(\mathbf{x}^{k-1})}(\mathbf{x})$ for each $k = 1, 2, \ldots$ and that $\left|\left|\mathbf{x}^k - \mathbf{x}^{k-1}\right|\right| \to 0$. Then $\{\mathbf{x}^k\}$ converges to a KKT point of the problem* (4.14)*, provided that point is in $\mathfrak{F}_0$ and satisfies the LICQ.*

Proof. It follows from the asymptotic slack control (2.9), from $\left|\left|\mathbf{x}^k - \mathbf{x}^{k-1}\right|\right| \to 0$, and from the fact that x^{k-1} must eventually become strictly timing feasible that we can write $\breve{N}^k_{\mathbf{d}(\mathbf{x}^{k-1})}(\mathbf{x})$ as $\sum_i(1+\theta_i^k)h_i(\mathbf{x}) + \sum_r \mu^k \breve{p}_r(D_r(\mathbf{x}))$, where $\theta_i^k \le \theta^k$ for all nets i and $\theta^k \to 0$. Now using analogous variables and Lagrange multiplier estimates to those given above, we get $\tilde{a}_m(\mathbf{d}(\mathbf{x}))$ and $\tilde{\tau}_j^k(\mathbf{x})$ as defined in in (4.6) and (4.19), respectively, and $\tilde{\lambda}_e^k(\mathbf{x}) = 0$, $\tilde{u}_j^k(\mathbf{x}) = 0$, and $\tilde{v}_\ell^k(\mathbf{x}) = 0$. Using these Lagrange multiplier estimates, it is straightforward to verify that all KKT conditions are satisfied in the limit. □

The practical implication of Theorem 4.5 is that, given a placer that can find approximate local minimizers for the unconstrained subproblems $\min_{\mathbf{x}} N^k(\mathbf{x})$, for $k = 1, 2, \ldots$, we can use the previous placement $\mathbf{x}^{k-1}$ as initial guess for the next unconstrained local minimization. Then the placements $\mathbf{x}^k$ will converge to a very likely local minimum of the constrained problem (4.14). Note that we must ensure that the placements eventually become strictly timing feasible.

5 Implementation

Motivated by the results in Sections 3 and 4, to solve the problem (4.14), we can iteratively perform unconstrained minimization of $N^k_{\mathbf{d}(\mathbf{x}^{k-1})}(\mathbf{x})$ using the previous solution $\mathbf{x} \leftarrow \mathbf{x}^{k-1}$ as a starting point. In each outer iteration, we set $\mathbf{x}^k \leftarrow \mathbf{x}$ if $\mathbf{x}$ approximates a local minimizer within some tolerance τ^k, where $\tau^k \searrow 0$. This framework is described in Algorithm 5.1.

Algorithm 5.1 Example, iterative placement with net weighting.

```
Choose initial placement x^0, tolerances {τ^k} such that τ^k ↘ 0, and convergence tolerance ρ > 0.
k ← 0
while k ≤ 1 or ||x^k − x^{k−1}|| > ρ do
  k ← k + 1
  x ← x^{k−1}
  while ||∇N^k_{d(x^{k−1})}(x)|| > τ^k do
    iterate x in unconstrained minimization of N^k_{d(x^{k−1})}(x)
  end while
  x^k ← x
end while
```

We implement each of the weighted objectives described in Section 2.3 in the state-of-the-art placer mPL [4]. Although significantly more sophisticated than the example given in Algorithm 5.1, similar principles are followed in mPL. We update the weighting parameter once every outer loop iteration, using the previous iterate to calculate net weights, then solve the inner loop subproblems using the Uzawa algorithm [1] for smoothed density equality constraints. A decreasing schedule of tolerances for cell overlap is used to determine sufficient convergence of the iterates. We use the log-sum-exp wire length approximation [15, 9] and a quadratic delay model, which satisfy all conditions for h_i and d_e required in this framework.

6 Experimental Results

In Table 6.1, we measure the efficacy of each method implemented in mPL on selected MCNC benchmarks [18]. These circuits were used because they contained the necessary timing information; the ISPD '05 and '06 contest examples could not be used due to lack of such information. The Cadence RTL compiler was used to synthesize the circuits with the Nangate 45nm open cell library. The results represent the best result obtained for each scheme in 8 runs, measured by shortest max delay after

Table 6.1: Comparison of weighting schemes in mPL.

		Global Placement		Detailed Placement		
Circuit	Weight	HPWL	Delay	HPWL	Delay	CPU
ex5p	None	1 (1.61E+07)*	1 (27.87)	1 (1.65E+07)	1 (28.41)	1 (5.9)
	VPR-c	1.14	0.86	1.04	0.88	1.61
	PATH-c	1.07	0.95	1.01	0.93	2.16
	APlace-c	1.13	0.88	1.03	0.89	3.81
alu4	None	1 (1.98E+07)	1 (29.55)	1 (2.00E+07)	1 (29.63)	1 (7.3)
	VPR-c	1.02	0.82	1.02	0.85	1.31
	PATH-c	1.19	0.92	1.10	0.93	2.16
	APlace-c	1.02	0.95	1.01	0.95	1.93
apex2	None	1 (2.57E+07)	1 (29.91)	1 (2.61E+07)	1 (30.13)	1 (9.3)
	VPR-c	1.05	0.85	1.02	0.89	1.49
	PATH-c	1.17	0.98	1.12	0.96	1.77
	APlace-c	1.12	0.83	1.03	0.91	2.26
pdc	None	1 (11.3E+07)	1 (54.56)	1 (11.9E+08)	1 (55.35)	1 (27.1)
	VPR-c	1.09	0.80	1.03	0.82	1.48
	PATH-c	1.11	0.89	1.06	0.88	1.81
	APlace-c	1.08	0.78	1.03	0.82	7.22
apex4	None	1 (2.60E+07)	1 (29.80)	1 (2.64E+07)	1 (28.58)	1 (7.4)
	VPR-c	1.04	0.87	1.01	0.95	1.42
	PATH-c	1.10	0.88	1.06	0.91	1.72
	APlace-c	1.02	0.88	1.00	0.97	2.47
des	None	1 (5.55E+07)	1 (32.59)	1 (5.60E+07)	1 (32.52)	1 (17.7)
	VPR-c	1.02	0.91	1.00	0.93	1.29
	PATH-c	1.05	0.95	1.03	0.98	1.48
	APlace-c	1.01	0.92	1.00	0.93	2.64
ex1010	None	1 (2.79E+07)	1 (32.25)	1 (2.83E+07)	1 (32.27)	1 (8.6)
	VPR-c	1.07	0.82	1.03	0.87	1.34
	PATH-c	1.08	0.84	1.06	0.85	1.72
	APlace-c	1.05	0.86	1.02	0.89	2.52
misex3	None	1 (1.74E+07)	1 (27.36)	1 (1.78E+07)	1 (26.22)	1 (6.9)
	VPR-c	1.07	0.85	1.02	0.95	1.42
	PATH-c	1.05	0.92	1.03	0.96	1.63
	APlace-c	1.08	0.82	1.03	0.93	2.09
seq	None	1 (3.25E+07)	1 (30.34)	1 (3.28E+07)	1 (30.53)	1 (10.2)
	VPR-c	1.04	0.82	1.03	0.84	1.32
	PATH-c	1.11	0.84	1.08	0.87	1.55
	APlace-c	1.05	0.81	1.03	0.86	1.83
spla	None	1 (1.11E+08)	1 (53.84)	1 (1.11E+08)	1 (53.33)	1 (23.6)
	VPR-c	1.01	0.84	1.01	0.90	1.32
	PATH-c	1.16	0.84	1.13	0.85	1.70
	APlace-c	1.04	0.85	1.01	0.91	5.17
Average	None	1.00	1.00	1.00	1.00	1.00
	VPR-c	1.05	0.85	1.02	0.89	1.40
	PATH-c	1.11	0.90	1.07	0.91	1.77
	APlace-c	1.06	0.86	1.02	0.91	3.19

*units are in microns

detailed placement; statistics are given for the placement after both global and detailed placement. The net weights were updated once every outer iteration in the mPL placer, which occurred approximately 50-70 times before sufficient convergence was attained. Note that net weightings were only applied during the global placement phase, so some degradation in the delays can be observed after the global placement phase. Column "HPWL" gives the half-perimeter wire length, "Delay" gives the max path delay of the circuit, and "CPU" is a measure of the computation time necessary to complete the placement. The values are scaled against those obtained using the regular, non-weighted mPL placer. On average, the VPR-c scheme yields the best improvement in delay, the smallest increase in CPU time compared to non-weighted mPL, and matches the APlace-c weights in smallest increase in wire length compared to non-weighted mPL.

In Table 6.2, we compare the VPR-c weighting method against the original method in [11], which we term "VPR." The difference between the two schemes is that the VPR scheme is flat, without increasing net weighting parameter, and the max delay is set dynamically to be the current max delay at every static timing analysis (VPR). The entries in the table are arranged as those in Table 6.1, with the exception that computation time has been omitted, as it did not vary significantly between the two methods (static timing analysis and re-calculation of net weights are carried out at every outer iteration in both methods). As in Table 6.1, the best result over 8 runs for each method is shown, measured by the shortest max delay after detailed placement. After detailed placement, the VPR-c scheme nearly matched the VPR scheme in wire length while outperforming it in delay on all 10 benchmarks, 4% on average. Thus, the modifications necessary for theoretical convergence yield

Table 6.2: Comparison of VPR-like methods in mPL.

		Global Placement		Detailed Placement	
Circuit	Weight	HPWL	Delay	HPWL	Delay
ex5p	VPR-c	1.04	0.90	1.01	0.90
	VPR	1.01	0.93	1.01	0.93
alu4	VPR-c	1.02	0.82	1.02	0.85
	VPR	1.01	0.91	1.02	0.92
apex2	VPR-c	1.05	0.85	1.02	0.89
	VPR	1.01	0.95	1.02	0.95
pdc	VPR-c	1.09	0.80	1.03	0.82
	VPR	1.04	0.87	1.01	0.87
apex4	VPR-c	1.04	0.87	1.01	0.95
	VPR	1.01	0.87	1.01	0.96
des	VPR-c	1.02	0.91	1.00	0.93
	VPR	1.03	0.91	1.02	0.94
ex1010	VPR-c	1.07	0.82	1.03	0.87
	VPR	1.01	0.88	1.01	0.91
misex3	VPR-c	1.07	0.85	1.02	0.95
	VPR	1.01	0.89	1.01	0.97
seq	VPR-c	1.04	0.82	1.03	0.84
	VPR	1.01	0.89	1.01	0.89
spla	VPR-c	1.01	0.84	1.01	0.90
	VPR	1.00	0.92	1.00	0.91
Average	VPR-c	1.05	0.85	1.02	0.89
	VPR	1.01	0.90	1.01	0.93

an improvement over the method in [11].

7 Conclusions and Future Work

In this work, we determined a set of criteria defining a class of net weighting schemes that were shown to converge to optimal placements for the original timing-constrained problem. When a *global* minimizer to the unconstrained weighted objective can be found, the essential property for convergence of a weighting algorithm is the *asymptotic slack control.* In the more practical case when only an *approximate local* minimizer to the unconstrained weighted objective can be found, a weighting must also be *nondecreasing, non-critically indifferent,* and *path consistent.* Several schemes satisfying these properties were identified and implemented in the mPL placer on the MCNC benchmarks. The *VPR-c* scheme outperformed all others, improving delay by an average of 11% while increasing wire length by 2% and increasing computation time by 40%, as compared to unweighted placement. Further, VPR-c outperformed the original VPR weighting scheme with improved delay in all MCNC benchmarks tested, an average of 4%; additionally, minimal increase in wire length and no change in computation time was observed.

In the future, we plan to implement Lagrangian schemes for comparison with our net weighting methodology. It can be shown that, using a scheme with projection of the multipliers as in [6], the method can be viewed as a net weighting that satisfies many of the required properties in the framework. We would like to further investigate this and the potential insights it can provide. In addition, we plan to extend the net weighting methodology to handle the timing-driven sizing and placement problem.

8 Acknowledgments

The authors gratefully acknowledge Guojie Luo and Kirill Minkovich of the UCLA Computer Science Department, and John Lee and Prof. Lieven Vandenberghe of the UCLA Electrical Engineering Department.

A Notation

A summary of the notation used throughout this paper is provided for reference in Table A.1.

Table A.1: Notation

Symbol	Definition
$\mathbf{x}$	The vector of pin locations of the circuit
T	The desired upper timing bound for the circuit
$\mathcal{N}$	The set of all nets in the circuit
$\mathcal{P}$	The (finite) set of all pins in the circuit
$\mathcal{P}_{\mathcal{I}}$	The set of all input pins in the circuit
$\mathcal{P}_{\mathcal{O}}$	The set of all output pins in the circuit
$\mathcal{P}_{\mathcal{M}}$	$\mathcal{P} \setminus (\mathcal{P}_{\mathcal{I}} \cup \mathcal{P}_{\mathcal{O}})$
Π	The set of all paths in the circuit
Π_e	The set of all paths containing edge e
Π_i	The set of all paths passing through net i
$\mathcal{E}$	The set of all edges in the circuit
S	$\|\Pi\|$, i.e. the number of paths in the circuit
M	$\|\mathcal{E}\|$, i.e. the number of edges in the circuit
$h_i(\mathbf{x})$	Function measuring estimated wire length of net i (see Sec. 2.1)
$d_e(\mathbf{x})$	Function measuring delay of edge e (see Sec. 2.1)
$\mathbf{d}(\mathbf{x})$	$[d_1(\mathbf{x}) \ \ldots \ d_M(\mathbf{x})]^T$
$w_i^k(\mathbf{d}(\mathbf{x}))$	Weighting function for net i and index k (see Sec. 2.2)
$G(\mathbf{x})$	$\sum_{i\in\mathcal{N}} h_i(\mathbf{x})$
$\Psi^k(\mathbf{x})$	$\sum_{i\in\mathcal{N}} w_i^k(\mathbf{d}(\mathbf{x}))h_i(\mathbf{x})$
$N^k(\mathbf{x})$	$G(\mathbf{x}) + \Psi^k(\mathbf{x})$
$\sigma_\pi(\mathbf{d}(\mathbf{x}))$	The slack of path π (see (2.3))
$A_t(\mathbf{d}(\mathbf{x}))$	Arrival time of pin t (see (2.4))
$R_s(\mathbf{d}(\mathbf{x}))$	Required arrival time of pin s (see (2.5))
$\sigma_e(\mathbf{d}(\mathbf{x}))$	The slack of edge e (see (2.6))
$\sigma_i(\mathbf{d}(\mathbf{x}))$	The slack of net i, defined as $\min_{e\in i}\sigma_e(\mathbf{d}(\mathbf{x}))$
$\mathfrak{F}$	$\{\mathbf{x} \| \ \sigma_\pi(\mathbf{d}(\mathbf{x})) \geq 0 \ \forall \text{ paths } \pi \in \Pi\}$
$\mathfrak{F}_0$	$\{\mathbf{x} \| \ \sigma_\pi(\mathbf{d}(\mathbf{x})) > 0 \ \forall \text{ paths } \pi \in \Pi\}$
$D_r(\mathbf{x})$	Generalized density constraint (see Sec. 3.2)
R	Number of generalized density constraints
$p_r^k(D_r(\mathbf{x}))$	Penalty function for generalized density constraint (see Sec. 3.2)
$\hat{N}^k(\mathbf{x})$	$G(\mathbf{x}) + \Psi^k(\mathbf{x}) + \sum_{r=1}^R p_r^k(D_r(\mathbf{x}))$
$\mathfrak{D}$	$\{\mathbf{x} : D_r(\mathbf{x}) = 0, \ r = 1, \ldots, R\}$
κ	Lower bound for $h_i(\mathbf{x})$ (see Sec. 4.1.1)
ξ	"Jaggedness" bounding constant for $N^k(\mathbf{x})$ (see Sec. 4.1.2)
a_j	Variable in (4.4) and (4.14); upper bound on the arrival time of pin j
$\mathbf{a}$	$(a_1, \ldots, a_n)$
$\mathcal{L}(\mathbf{x}, \mathbf{a}, \mathbf{u}, \mathbf{v}, \boldsymbol{\lambda})$	Lagrangian for the problem (4.4)
$\mathbf{u}$, $\mathbf{v}$, $\boldsymbol{\lambda}$	Lagrange multipliers corresponding to timing constraints
$\tilde{\mathbf{a}}(\mathbf{d}(\mathbf{x}))$	Variable $\mathbf{a}$ based on a placement $\mathbf{x}$ (see (4.6))
$\tilde{\mathbf{u}}^k(\mathbf{x})$, $\tilde{\mathbf{v}}^k(\mathbf{x})$, $\tilde{\boldsymbol{\lambda}}^k(\mathbf{x})$	Lagrange multiplier estimates based on a placement $\mathbf{x}$ (see (4.6))
$\check{\mathcal{L}}(\mathbf{x}, \mathbf{a}, \mathbf{u}, \mathbf{v}, \boldsymbol{\lambda}, \boldsymbol{\tau})$	Lagrangian for the problem (4.14)
$\boldsymbol{\tau}$	Lagrange multipliers corresponding to density constraints
$\tilde{\tau}_j^k(\mathbf{x})$	Lagrange multiplier estimate for τ_j (See 4.19)
$\check{N}^k(\mathbf{x})$	$G(\mathbf{x}) + \Psi^k(\mathbf{x}) + \sum_{r=1}^R \mu^k \breve{p}_r(D_r(\mathbf{x}))$
$\breve{p}_r^k$	Penalty function for generalized density constraint (see Sec. 4.3)
μ^k	Penalty coefficient in $\breve{p}_r^k$, i.e. $\breve{p}_r^k = \mu^k \breve{p}_r$ (see Sec. 4.3)
$\breve{p}_r$	Index-independent component of penalty function $\breve{p}_r^k$ (see Sec. 4.3)
$\check{N}_{\mathbf{d}(\mathbf{y})}^k(\mathbf{x})$	$G(\mathbf{x}) + \sum_{i\in\mathcal{N}} w_i^k(\mathbf{d}(\mathbf{y}))h_i(\mathbf{x}) + \sum_{r=1}^R \mu^k \breve{p}_r(D_r(\mathbf{x}))$

References

[1] K. Arrow, L. Huriwcz and H. Uzawa, *Studies in Nonlinear Programming*, Stanford University Press, Stanford, CA, 1958.

[2] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.

[3] T. F. Chan, J. Cong, J. R. Shinnerl, K. Sze, and M. Xie, "mPL6: enhanced multilevel mixed-size placement," *Proc. 2006 Int'l Symp. on Phys. Design*, pp. 212-214, 2006.

[4] T. F. Chan, J. Cong, and K. Sze, "Multilevel generalized force-directed method for circuit placement," *Proc. 2005 Int'l Symp. on Phys. Design*, pp. 185-192, April 2005.

[5] T. F. Chan, J. Cong, and E. Radke, "A Rigorous Framework for Convergent Net Weighting Schemes in Timing-Driven Placement," *Proc. 2009 Int'l Conf. on Computer Aided Design*, November 2009.

[6] C.-P. Chen, C. C. N. Chu, and D. F. Wong, "Fast and exact simultaneous gate and wire sizing by Lagrangian relaxation," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 18, No. 7, pp. 1014-1025, July 1999.

[7] J. Cong , T. Kong , J. Shinnerl , M. Xie, and X. Yuan, "Large-scale circuit placement," *ACM Trans. Des. Automat. Electron. Syst.*, vol. 10, no. 2, pp. 389-430, Apr. 2005.

[8] H. Eisenmann and F. M. Johannes, "Generic global placement and floorplanning," *Proc. 35th ACM/IEEE Design Automation Conference*, pp. 269-274, 1998.

[9] A. B. Kahng and Q. Wang, "Implementation and extensibility of an analytic placer," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 24, No. 5, pp. 1-14, May 2005.

[10] T. Kong, "A novel net weighting algorithm for timing-driven placement," *Proc. 2002 IEEE/ACM Int'l Conf. on Computer-aided Design*, pp. 172-176.

[11] A. Marquardt, V. Betz, and J. Rose, "Timing-driven placement for FPGAs," *Proc. 2000 ACM/SIGDA Eighth Int'l Symp. on Field Programmable Gate Arrays*, pp. 203-213, February 2000.

[12] G.-J. Nam and J. Cong (eds.), *Modern Circuit Placement: Best Practices and Results*, Springer, 2007.

[13] S. G. Nash and A. Sofer. *Linear and Nonlinear Programming*, McGraw-Hill, 1996.

[14] J. Nocedal and S. Wright, *Numerical Optimization*, Springer, New York, NY, 1999, pp. 490-525.

[15] W. Naylor, R. Donelly, and L. Sha, "Non-linear optimization system and method for wire length and delay optimization for an automatic electronic circuit placer," US Patent 6671859, 2003.

[16] R. S. Tsay and J. Koehl, "An analytic net weighting approach for performance optimization in circuit placement," *Proc. Design Automation Conf.*, pp. 620-625, 1991.

[17] M. H. Wright, "Interior methods for constrained optimization," *Acta Numerica 1992*, pp. 341-407.

[18] S. Yang, "Logic synthesis and optimization benchmarks, version 3.0," tech. report, Microelectronics Center of North Carolina, Research Triangle Park, NC, 1991.