The Swap, Expansion and Exwap Moves – A Simple Derivation and Implementation

Ido Leichter

Computer Science Department
Technion - Israel Institute of Technology
Haifa 32000, Israel
idol@cs.technion.ac.il

June 30, 2008

Abstract. In the context of multi-label energy function minimization, the swap and the expansion are two types of moves that were introduced by Boykov et al. [3]. They proposed efficient algorithms for finding local minima with respect to each of these two moves when each energy term depends on two variables at most. The minimization was carried out by a sequence of optimal moves that were calculated by seeking minimum cuts of specially constructed weighted graphs. In this paper these optimal swap and expansion moves are obtained in a short and simple manner by incorporating the original algorithm by Greig et al. [5] as a “black box.” Our alternative derivation has three advantages over the original one: 1. Given Greig et al.’s original solution as a black box, it is shorter, purely algebraic (that is, no graphs are involved) and, we believe, simpler to understand and implement. 2. It is derived under more general conditions. 3. It contains a proof that the found local minima with respect to expansion moves are actually also local minima with respect to swap moves – a point that seems to have been overlooked in previous work. All the results are extended for energy terms that depend on three variables by using Kolmogorov et al.’s binary minimization algorithm [9] as a black box. In addition, the exwap move type, a generalization of the expansion and the swap move types, is introduced and an efficient algorithm for minimizing with respect to it is derived.

Index Terms
Energy minimization, minimum cut, maximum flow, Markov Random Fields.

1 Introduction

The problem discussed in this paper is the minimization of energy functions $E(x)$ where the variables $x = (x_1, \ldots, x_n)$ may assume labels from the label set $L = \{l_0, l_1, \ldots, l_{L-1}\}$. In the beginning the concerned energy functions will have the structure

$$E(x) = \sum_{k=1}^{n} E_k(x_k) + \sum_{1 \leq i < j \leq n} E_{ij}(x_i, x_j),$$

and all the results will be derived for it. Afterward, the results are generalized for energy functions that contain terms that are functions of three variables as well. Under the restricted case where the labels have a linear ordering and the $E_{ij}$ terms are convex functions of the
difference of the labels’ ordinal numbers, the energy function (1) can minimized fast and exactly [6]. Generally, however, minimizing (1) is NP-hard already for \( L = 2 \). This is easily seen from the fact that MAX-2-SAT is NP-hard (see, for example, [1]) and has an immediate reduction to the former problem.

Given a labeling \( x \in L^n \) and a pair of labels \( \alpha \) and \( \beta \), an \( \alpha \)-\( \beta \)-swap is any move (that is, changes applied on the labeling) that consists of only alterations of variables between \( \alpha \) and \( \beta \). Given a labeling and a label \( \alpha \), an \( \alpha \)-expansion is any move that consists of only alterations of variables to \( \alpha \). These two types of moves were introduced by Boykov et al. in [3] (and earlier in [2]), where they developed an efficient scheme for finding the \( \alpha \)-\( \beta \)-swap that yields the highest reduction of the energy function (1) out of all \( \alpha \)-\( \beta \)-swaps, and similarly for the expansion move type. By repeatedly performing such optimal moves of one type, each time for a different label (for expansion moves) or label pair (for swap moves) in a fixed or random order and until no reduction of the energy function is possible for any move of the corresponding type, the energy function is minimized with respect to this type of move. These algorithms proved to yield good results for many problems (e.g., [10]).

The schemes in [3] for finding the optimal swap and expansion moves were based on seeking minimum cuts of specially constructed weighted graphs. The scheme for finding the optimal swap move was developed under the condition that each \( E_{ij} \) term is a semi-metric, that is, for any labels \( \alpha, \beta \in L \) this term satisfies

\[
E_{ij}(\alpha, \beta) = 0 \iff \alpha = \beta, \tag{2}
\]

\[
E_{ij}(\alpha, \beta) = E_{ij}(\beta, \alpha) \geq 0. \tag{3}
\]

The scheme for finding the optimal expansion move was developed under the condition that each \( E_{ij} \) term is a metric, that is, it is a semi-metric and additionally satisfies

\[
E_{ij}(\alpha, \beta) \leq E_{ij}(\alpha, \gamma) + E_{ij}(\gamma, \beta) \tag{4}
\]

for any labels \( \alpha, \beta, \gamma \in L \). As noted in [3], the symmetry condition in the equality part of (3) can be eliminated by using more complicated schemes that require the use of directed graphs.

In 1989 Greig et al. provided in [5] an efficient algorithm for finding, under a certain condition, the maximum of a quadratic function of binary variables by reducing the maximization problem to that of finding the minimum cut/maximum flow in a flow network. It is straightforward to see that this maximization is equivalent to the minimization of energy functions of the structure (1) for \( L = 2 \) under a certain corresponding condition (see Sec. 2). We believe that Greig et al.’s algorithm for the minimization of energy functions was the first one in computer vision that was based on graph cuts.

This paper proposes an alternative derivation of the optimal swap and expansion moves for the energy function (1) by using Greig et al.’s original solution. By using it as a “black box”, the derivation given in this paper is shorter and purely algebraic (that is, no graphs are used outside this black box), which makes it simpler to understand and simpler to implement. Moreover, the conditions under which the solution here is derived are more general than those in [3]. In addition, the proposed derivation includes a proof that under the swap move-related condition (which, as will be explained, is more general than the expansion move-related one) a local minimum with respect to expansion moves is actually also a local minimum with respect to swap moves. All these results are extended to the case of energy functions consisting of terms that are functions of up to three variables. These extensions remain simple and purely algebraic via the use, as a black box as well, of Kolmogorov et al.’s algorithm in [9] for minimizing such energy functions of binary variables. In addition, we introduce the exwap
move type – a generalization of both the expansion and the swap move types. An efficient algorithm for minimizing the extended energy function with respect to this move is derived.

A recent related work is [7], where algorithms for finding the optimal expansion and swap moves were proposed for energy functions consisting of terms that are functions of an arbitrary number of variables. However, for energy functions whose terms are functions of three variables at most, the conditions assumed there are stricter than the corresponding ones here. For example, dealing with the $P^3$ Potts model [7] required a separate treatment that included a special graph construction. The algorithm proposed in this paper may be applied in the $P^3$ Potts model as is and without any graph constructions (outside the black box consisting of the aforementioned Kolmogorov et al.’s algorithm). Like [3], [7] does not include any results regarding the relation between the two types of local minima.

The paper proceeds as follows: Sec. 2 states the original problem solved by Greig et al. in [5] and shows its equivalence to the minimization of energy functions (1) for $L = 2$. Derivations of the optimal swap and expansion moves for energy functions (1) where $L > 2$ are given in Sec. 3 and Sec. 4, respectively. Sec. 5 shows for these energy functions that local minima with respect to expansion moves are also local minima with respect to swap moves, provided that the swap move-related condition is met. Sec. 6 extends the above results to energy functions consisting of terms that are functions of up to three variables. Sec. 7 introduces the exwap move type and derives the algorithm for the minimization with respect to it. A conclusion is given in Sec. 8.

2 The Original Solution to the Binary Case

Given $n$ binary variables, $x = (x_1, x_2, \ldots, x_n)$, $x_i \in \{0, 1\}$, Greig et al. have given in [5] an efficient solution for calculating

$$\arg \max_x \sum_{k=1}^n \lambda_k x_k + \frac{1}{2} \sum_{i,j = 1, i \neq j}^n \beta_{ij} \{x_i x_j + (1 - x_i)(1 - x_j)\}, \quad \beta_{ij} = \beta_{ji} \geq 0. \quad (5)$$

The solution is based on a reduction of the problem to that of finding the minimum cut/maximum flow in a flow network. Opening the parentheses in the objective function and grouping common terms bring it to the canonical form of a quadratic function. Negating this quadratic function and eliminating the constant term results in the equivalent optimization problem

$$\arg \min_x \sum_{k=1}^n \lambda'_k x_k + \sum_{1 \leq i < j \leq n} \beta'_{ij} x_i x_j, \quad \beta'_{ij} \leq 0. \quad (6)$$

The original coefficients in (5) are obtained from those in (6) by the simple relations

$$\beta_{ij} = \beta'_{ji},$$
$$\lambda_k = -\lambda'_k - 0.5 \sum_{j=1}^n \beta'_{kj}. \quad (7)$$

The fact that (6) may be solved efficiently via graph cut techniques was also proved in [4], but without relying on the original solution to (5). This has made the proof there more complicated and involve a graph construction.

The solution to (6) may be easily utilized for minimizing the energy function (1) for $L = 2$ by assigning $l_0 = 0$ and $l_1 = 1$ and expressing each energy term as a function of its
0-1 variables:

\[
\arg\min_x \sum_{k=1}^{n} E_k(x_k) + \sum_{1 \leq i < j \leq n} E_{ij}(x_i, x_j)
\]

\[
= \arg\min_x \sum_{k=1}^{n} a_k x_k + \sum_{1 \leq i < j \leq n} \{b_{ij} x_i + c_{ij} x_j + d_{ij} x_i x_j\}, \quad (8)
\]

where

\[
a_k = E_k(l_1) - E_k(l_0),
\]

\[
b_{ij} = E_{ij}(l_1, l_0) - E_{ij}(l_0, l_0),
\]

\[
c_{ij} = E_{ij}(l_0, l_1) - E_{ij}(l_0, l_0),
\]

\[
d_{ij} = E_{ij}(l_1, l_1) - E_{ij}(l_1, l_0) - E_{ij}(l_0, l_1) + E_{ij}(l_0, l_0).
\]

(Note that the additive constants in all terms were eliminated as they are irrelevant in the optimization.) Opening the parentheses and grouping common terms results in an equivalent optimization problem in the form of (6) with the coefficients

\[
\lambda'_k = a_k + \sum_{i=1}^{k-1} b_{ik} + \sum_{j=k+1}^{n} c_{kj}, \quad (10)
\]

\[
\beta'_{ij} = d_{ij}. \quad (11)
\]

Since (6) is restricted by \(\beta'_{ij} \leq 0\), we obtain that the energy can be minimized using Greig et al.’s solution to (5) as long as \(d_{ij} \leq 0\), that is, as long as

\[
E_{ij}(l_1, l_1) + E_{ij}(l_0, l_0) \leq E_{ij}(l_1, l_0) + E_{ij}(l_0, l_1).
\]

This is the regularity condition in [9], where it was derived without using the original solution to (5), resulting in a derivation that was more complicated and involved a graph construction.

3 Finding the Optimal Swap Move

Denote the current labeling (for which the optimal \(\alpha\)-\(\beta\)-swap is sought) by \(\{x_i^0\}_{i=1}^{n}\) and the index-set of all variables currently labeled \(\alpha\) or \(\beta\) by \(I = \{i : x_i^0 \in \{\alpha, \beta\}\}\). The problem of finding the optimal \(\alpha\)-\(\beta\)-swap is

\[
\arg\min_{x_i \in \{\alpha, \beta\}} \min_{i \in I} E\left(\{x_i\}_{i \in I}, \{x_i^0\}_{i \notin I}\right)
\]

\[
= \arg\min_{x_i \in \{\alpha, \beta\}} \sum_{k \in I} E_k(x_k) + \sum_{k \in I} E_k(x_k^0) + \sum_{1 \leq i < j \leq n} 1 \in I, \ j \notin I
\]

\[
E_{ij}(x_i, x_j) + \sum_{1 \leq i < j \leq n} 1 \notin I, \ j \notin I
\]

\[
E_{ij}(x_i^0, x_j^0), \quad (13)
\]

Eliminating the second and last sums, which are constants, and grouping common terms results in the equivalent minimization problem

$$\arg \min_{x_i \in \{\alpha, \beta\}} \sum_{k \in I} E'_k(x_k) + \sum_{1 \leq i < j \leq n} E_{ij}(x_i, x_j),$$  \hspace{1cm} (14)

where

$$E'_k(x_k) = E_k(x_k) + \sum_{1 \leq i < k} E_{ik}(x^0_i, x_k) + \sum_{k < j \leq n} E_{kj}(x_k, x^0_j). \hspace{1cm} (15)$$

The objective function is of the same structure as the one of the energy function (1) for \(L = 2\) (associate \(\alpha\) with \(l_0\) and \(\beta\) with \(l_1\), or vice versa). Therefore, it may be minimized using Greig et al.’s solution in the manner described in Sec. 2, provided that each \(E_{ij}\) term fulfills (12) for labels \(\alpha\) and \(\beta\). Since the overall minimization procedure iterates over all label pairs, this condition should hold for all of them, that is,

$$\forall \alpha, \beta \in \mathcal{L} \quad E_{ij}(\alpha, \alpha) + E_{ij}(\beta, \beta) \leq E_{ij}(\alpha, \beta) + E_{ij}(\beta, \alpha). \hspace{1cm} (16)$$

Obviously, this condition is a relaxation of the conditions (2)-(3) assumed in [3], even when the symmetry condition is eliminated.

4 Finding the Optimal Expansion Move

As before, denote the current labeling by \(\{x^0_i\}_{i=1}^n\). The problem of finding the optimal \(\alpha\)-expansion is

$$\arg \min_{x_i \in \{\alpha, x^0_i\}} E(\{x_i\}_{i=1}^n) = \arg \min_{x_i \in \{\alpha, x^0_i\}} \sum_{k=1}^n E_k(x_k) + \sum_{1 \leq i < j \leq n} E_{ij}(x_i, x_j). \hspace{1cm} (17)$$

As in the previous case, the objective function has the same structure as the one of the energy function (1) for \(L = 2\) (for each variable \(x_i\), associate \(\alpha\) with \(l_0\) and \(x^0_i\) with \(l_1\), or vice versa). Therefore, it may be minimized using Greig et al.’s solution (Sec. 2) as well, provided that each \(E_{ij}\) term fulfills (12) for the labels feasible by an \(\alpha\)-expansion from the current labeling. For each \(x_i\), these feasible labels are \(\{\alpha, x^0_i\}\). Either associating \(\alpha\) with \(l_0\) for all variables or associating \(\alpha\) with \(l_1\) for all variables results in the following condition:

$$E_{ij}(\alpha, \alpha) + E_{ij}(x^0_i, x^0_j) \leq E_{ij}(\alpha, x^0_j) + E_{ij}(x^0_i, x^0_j). \hspace{1cm} (18)$$

Generally \(x^0_i\) and \(x^0_j\) might equal any labels \(\beta\) and \(\gamma\), and the overall minimization procedure iterates over all labels \(\alpha\). Therefore, the above condition should be fulfilled for all label triplets, that is,

$$\forall \alpha, \beta, \gamma \in \mathcal{L} \quad E_{ij}(\alpha, \alpha) + E_{ij}(\beta, \gamma) \leq E_{ij}(\alpha, \gamma) + E_{ij}(\beta, \alpha). \hspace{1cm} (19)$$

It is straightforward to verify that this condition is a relaxation of the conditions (2)-(4) assumed in [3], even when the symmetry condition is eliminated.

Substituting \(\beta\) for \(\gamma\) in (19) yields (16), which shows that the obtained swap move-related condition is a relaxation of the obtained expansion move-related condition.
5 Expansion-Related Minima Are Swap-Related Minima

Following we prove that under the swap move related condition (16) (which, as was explained, is more general than the expansion move related one), a local minimum of the energy function (1) with respect to expansion moves is also a local minimum with respect to swap moves.

Proof. Assume the labeling \( \{x_i^0\}_{i=1}^n \) is a local minimum with respect to all expansion moves. We will show that under condition (16) any \( \alpha \rightarrow \beta \)-swap move from this labeling will increase the energy function (1).

Denote the labeling after the \( \alpha \rightarrow \beta \)-swap move by \( \{x_i^+\}_{i=1}^n \), the alteration of variable \( x_i \) from label 0\(^{-}\) to label 1\(^{+}\) as part of an \( \alpha \rightarrow \beta \)-swap move by \( \overrightarrow{x_i^0} \rightarrow \overrightarrow{x_i^+} = 0^0 \rightarrow 1^+ \), and denote the set of all other possible alterations of this variable by \( \overrightarrow{x_i} \in 0^0 \rightarrow 1^+ \) (note that alterations of the type \( \overrightarrow{x_i} = l \rightarrow l \), that is, no alteration de facto, are legitimate). To shorten the formulas, we denote \( E_{ij}(x_i, x_j) \equiv E_{ij}(x_i, x_j) (i \neq j) \) and perform all the double summations over unordered pairs of variables. The difference between the energy function after an \( \alpha \rightarrow \beta \)-swap move and before the move is

\[
E \left( \{x_i^+\}_{i=1}^n \right) - E \left( \{x_i^0\}_{i=1}^n \right) = \sum_{k, \overrightarrow{x_k} = \alpha \rightarrow \beta} [E_k(\beta) - E_k(\alpha)] + \sum_{k, \overrightarrow{x_k} = \beta \rightarrow \alpha} [E_k(\alpha) - E_k(\beta)] + \sum_{\{i, j\}, i \neq j} \left[ E_{ij}(\beta, x_j^+) - E_{ij}(\beta, x_j^0) \right] + \sum_{\{i, j\}, i \neq j} \left[ E_{ij}(\alpha, x_j^+) - E_{ij}(\alpha, x_j^0) \right].
\]

By inequality (16) we have the following lower bound for the term inside the last sum:

\[
E_{ij}(\beta, \alpha) - E_{ij}(\alpha, \beta) \geq E_{ij}(\alpha, \alpha) + E_{ij}(\beta, \beta) - 2E_{ij}(\alpha, \beta).
\]

Substituting this lower bound for the term inside the last sum and splitting the resulting sum into two sums results in the following lower bound for the difference in the energy function:

\[
E \left( \{x_i^+\}_{i=1}^n \right) - E \left( \{x_i^0\}_{i=1}^n \right) \geq \sum_{k, \overrightarrow{x_k} = \alpha \rightarrow \beta} [E_k(\beta) - E_k(\alpha)] + \sum_{k, \overrightarrow{x_k} = \beta \rightarrow \alpha} [E_k(\alpha) - E_k(\beta)] + \sum_{\{i, j\}, i \neq j} \left[ E_{ij}(\beta, x_j^+) - E_{ij}(\beta, x_j^0) \right] + \sum_{\{i, j\}, i \neq j} \left[ E_{ij}(\alpha, x_j^+) - E_{ij}(\alpha, x_j^0) \right].
\]

The first, third and fifth sums comprise exactly the difference in the energy function resulting from making only the \( \alpha \rightarrow \beta \) alterations in the considered \( \alpha \rightarrow \beta \)-swap. These alterations are
equivalent to a $\beta$-expansion from the labeling $\{x_i^0\}_{i=1}^n$. The second, fourth and sixth sums comprise exactly the difference in the energy function resulting from making only the $\beta$-to-$\alpha$ alterations in the considered $\alpha$-$\beta$-swap. These alterations are equivalent to an $\alpha$-expansion from the labeling $\{x_i^0\}_{i=1}^n$. By assumption all expansion moves from the labeling $\{x_i^0\}_{i=1}^n$ do not reduce the energy function, and therefore the right-hand side of (22) is nonnegative.

6 Extensions for Terms Consisting of Three Variables

All the above results were derived for energy functions of the structure (1), which consist of terms that are functions of up to two variables. In the following is discussed the generalization of these results to energy functions that contain terms that are functions of three variables as well, that is, energy functions of the form

$$E(x) = \sum_{k=1}^{n} E_k(x_k) + \sum_{1 \leq i < j \leq n} E_{ij}(x_i, x_j) + \sum_{1 \leq i < j < k \leq n} E_{ijk}(x_i, x_j, x_k),$$

(23)

where, as before, the variables $x = (x_1, \ldots, x_n)$ may assume labels from the label set $L = \{l_0, l_1, \ldots, l_{L-1}\}$. In the following the energy function (23) will be referred as the extended energy function. The generalization is accomplished by replacing the algorithm component that minimizes the energy function (1) for $L = 2$ (Sec. 2) with Kolmogorov et al.’s algorithm in [9] for minimizing the energy function (23) for $L = 2$. The latter algorithm finds the global minimum under the condition that all the $E_{ij}$ terms satisfy constraint (12) as before, and that all six projections of two variables of each $E_{ijk}$ term satisfy this constraint as well.

As defined in [9], a projection of a function of binary variables is a function of a subset of these variables, where this function is obtained by fixing all the variables outside this subset.

6.1 The Optimal Swap Move for the Extended Energy Function

Using the notation that was used in the corresponding previous derivation (Sec. 3) and performing a similar derivation to the one performed there, but this time for the energy function (23), shows that the problem of finding the optimal $\alpha$-$\beta$-swap move is equivalent here to the minimization problem

$$\arg \min_{x_i \in \{\alpha, \beta\}} \sum_{i \in I} E''_k(x_k) + \sum_{1 \leq i < j \leq n} E''_{ij}(x_i, x_j) + \sum_{1 \leq i < j < k \leq n} E_{ijk}(x_i, x_j, x_k).$$

(24)

The terms $E''_k(x_k)$ and $E''_{ij}(x_i, x_j)$ are sums of projections of terms in (23) of the corresponding variables, where we generalize the notion of a projection of a function to include functions of variables that may assume more than two labels.

Actually, the condition in [9] is that all projections of the whole function $E(X)$ of two variables satisfy (12), which is a more general condition than the one assumed here (this follows from the regrouping theorem in [9]). Nevertheless, for simplicity we shall restrict the derivations to the stricter condition.
The above objective function is of the same structure as the one of the energy function (23) for \( L = 2 \) (as before, associate \( \alpha \) with \( l_0 \) and \( \beta \) with \( l_1 \), or vice versa). Therefore, if all the \( E_{ij} \) terms and all projections of all the \( E_{ijk} \) terms of two variables fulfill (12) for any pair of labels, the minimization algorithm in [9] can be used to solve the minimization problem (24). These conditions may be summarized as follows: for all terms that are functions of two or three variables and for all \( \alpha, \beta, \gamma \in \mathcal{L} \),

\[
E_{ij}(\alpha, \alpha) + E_{ij}(\beta, \beta) \leq E_{ij}(\alpha, \beta) + E_{ij}(\beta, \alpha), \\
E_{ijk}(\alpha, \alpha, \gamma) + E_{ijk}(\beta, \beta, \gamma) \leq E_{ijk}(\alpha, \beta, \gamma) + E_{ijk}(\beta, \alpha, \gamma), \\
E_{ijk}(\alpha, \gamma, \alpha) + E_{ijk}(\beta, \gamma, \beta) \leq E_{ijk}(\alpha, \gamma, \beta) + E_{ijk}(\beta, \alpha, \gamma), \\
E_{ijk}(\gamma, \alpha, \alpha) + E_{ijk}(\gamma, \beta, \beta) \leq E_{ijk}(\gamma, \alpha, \beta) + E_{ijk}(\gamma, \beta, \alpha).
\] (25)

### 6.2 The Optimal Expansion Move for the Extended Energy Function

As in the corresponding previous derivation (Sec. 4), denote the current labeling by \( \{x_i^0\}_{i=1}^n \). The problem of finding the optimal \( \alpha \)-expansion this time is

\[
\arg \min_{x_i \in \{\alpha, x_i^0\}} \sum_{k=1}^{n} E_k(x_k) + \sum_{1 \leq i < j \leq n} E_{ij}(x_i, x_j) + \sum_{1 \leq i < j < k \leq n} E_{ijk}(x_i, x_j, x_k). \quad (26)
\]

As in the previous case, the objective function has the same structure as the one of the energy function (23) for \( L = 2 \) (for each variable \( x_i \), associate \( \alpha \) with \( l_0 \) and \( x_i^0 \) with \( l_1 \), or vice versa). Therefore, it may be minimized using the minimization algorithm in [9], provided that all the \( E_{ij} \) terms and all the projections of the \( E_{ijk} \) terms of two variables fulfill (12), where the variables in (12) and the fixed variables in the projections may assume the labels feasible by an \( \alpha \)-expansion from the current labeling. For each \( x_i \), these feasible labels are \( \{\alpha, x_i^0\} \). As in the corresponding previous derivation (Sec. 4), we either associate \( \alpha \) with \( l_0 \) for all variables or associate \( \alpha \) with \( l_1 \) for all variables. Generally, \( x_i^0 \), \( x_j^0 \), and \( x_k^0 \) might equal any labels \( \beta, \gamma, \) and \( \delta \), respectively, and the overall minimization procedure iterates over all labels \( \alpha \). Therefore, the conditions under which the proposed minimization procedure works may be summarized as follows: for all terms that are functions of two or three variables and for all \( \alpha, \beta, \gamma, \delta \in \mathcal{L} \),

\[
E_{ij}(\alpha, \alpha) + E_{ij}(\beta, \beta) \leq E_{ij}(\alpha, \gamma) + E_{ij}(\beta, \alpha), \\
E_{ijk}(\alpha, \alpha, \delta) + E_{ijk}(\beta, \beta, \delta) \leq E_{ijk}(\alpha, \gamma, \delta) + E_{ijk}(\beta, \alpha, \delta), \\
E_{ijk}(\alpha, \delta, \alpha) + E_{ijk}(\beta, \delta, \beta) \leq E_{ijk}(\alpha, \delta, \gamma) + E_{ijk}(\beta, \delta, \alpha), \\
E_{ijk}(\delta, \alpha, \alpha) + E_{ijk}(\delta, \beta, \beta) \leq E_{ijk}(\delta, \alpha, \gamma) + E_{ijk}(\delta, \beta, \alpha).
\] (27)

Substituting \( \beta \) for \( \gamma \) and substituting \( \gamma \) for \( \delta \) in (27) yields (25), which shows that the obtained swap move-related conditions are a relaxation of the obtained expansion move-related conditions also for energy functions from the extended class.

### 6.3 Expansion-Related Minima are Swap-Related Minima in the Extended Energy Function

Given an energy function consisting of terms that are functions of up to two variables, we proved in Sec. 5 that under the swap move-related condition (16), a local minimum with respect to expansion moves is also a local minimum with respect to swap moves. In the following we extend the proof for energy functions that consist of terms that are functions of up to three variables. The proof is extended under the extended swap move-related conditions (25).
Proof Extension. Denote the right-hand side of (20) by $A$. The difference between the energy function after an $\alpha$-$\beta$-swap move and the move becomes

$$E\left(\left\{ x_i^+ \right\}_{i=1}^n \right) - E\left(\left\{ x_i^0 \right\}_{i=1}^n \right) = A + \sum_{\{i, j, k\}, \text{all distinct}} \left[ E_{ijk}(\beta, x_j^+, x_k^+) - E_{ijk}(\alpha, x_j^0, x_k^0) \right]$$

$$+ \sum_{\{i, j, k\}, \text{all distinct}} \left[ E_{ijk}(\alpha, x_j^+, x_k^+) - E_{ijk}(\beta, x_j^0, x_k^0) \right]$$

By inequality (25) we obtain the following lower bound for the term inside the second sum from the end:

$$E_{ijk}(\beta, \beta, \alpha) - E_{ijk}(\alpha, \alpha, \beta) \geq E_{ijk}(\beta, \alpha, \alpha) + E_{ijk}(\beta, \beta, \beta) - E_{ijk}(\beta, \alpha, \beta) - E_{ijk}(\alpha, \alpha, \beta) \geq E_{ijk}(\beta, \alpha, \beta) + E_{ijk}(\alpha, \alpha, \alpha) - E_{ijk}(\alpha, \beta, \beta) + E_{ijk}(\alpha, \alpha, \alpha) - E_{ijk}(\alpha, \alpha, \beta) - E_{ijk}(\alpha, \alpha, \beta)$$

$$= E_{ijk}(\alpha, \alpha, \alpha) + E_{ijk}(\beta, \beta, \beta) - 2E_{ijk}(\alpha, \alpha, \beta). \quad (29)$$

Similarly, we obtain the following lower bound for the term inside the last sum:

$$E_{ijk}(\beta, \alpha, \alpha) - E_{ijk}(\alpha, \beta, \beta) \geq E_{ijk}(\alpha, \alpha, \alpha) + E_{ijk}(\beta, \beta, \beta) - 2E_{ijk}(\alpha, \alpha, \beta). \quad (30)$$

Denote the sum of the right-hand side of (22) and the first two sums following $A$ in (28) by $B$. Substituting these two lower bounds for the corresponding terms inside the last two sums and splitting each of the resulting sums into two sums results in the following lower bound.
for the difference in the energy function:

\[
E \left( \{ x^+ \}_{i=1}^n \right) - E \left( \{ x^0 \}_{i=1}^n \right) \geq B + \sum_{\{i, j, k\}, \text{all distinct}} [E_{ijk}(\beta, \beta, \beta) - E_{ijk}(\alpha, \alpha, \beta)]
\]

\[+ \sum_{\{i, j, k\}, \text{all distinct}} [E_{ijk}(\alpha, \alpha, \alpha) - E_{ijk}(\alpha, \alpha, \beta)]
\]

\[+ \sum_{\{i, j, k\}, \text{all distinct}} [E_{ijk}(\beta, \beta, \beta) - E_{ijk}(\alpha, \beta, \beta)]
\]

\[+ \sum_{\{i, j, k\}, \text{all distinct}} [E_{ijk}(\alpha, \alpha, \alpha) - E_{ijk}(\alpha, \beta, \beta)]. \tag{31}
\]

Similarly to the argument regarding (22), the sums in (31) (including those in \(B\)) can be partitioned into those comprising the difference in the energy function resulting from a \(\beta\)-expansion (first, third, fifth,...,eleventh sums) and those comprising the difference for an \(\alpha\)-expansion (second, fourth, sixth,...,twelfth sums). By assumption all expansion moves from the labeling \(\{ x^0 \}_{i=1}^n \) do not reduce the energy function, and therefore the right-hand side of (31) is nonnegative.

7 The Exwap Move

A new type of large move that may alter the label of all the variables regardless of the current labeling is the exwap move type. Given an ordered pair of labels \((\alpha, \beta)\), we define an \(\alpha\)-\(\beta\)-exwap move to be the unification of an \(\alpha\)-expansion move and an \(\alpha\)-\(\beta\) swap move. That is, the label alterations allowed by this move are alterations from any label to \(\alpha\) and alterations from \(\alpha\) to \(\beta\). Note that this move type is a generalization of both the swap move and the expansion move types. Like in the previous two move types, minimization of an energy function with respect to this move type can be performed by repeatedly performing optimal moves of this type, each time for a different label pair and until no reduction of the energy function is possible for any move of this type.

The method for finding the optimal \(\alpha\)-\(\beta\)-exwap move for the extended energy function (23) will now be derived. As before, denote the current labeling by \(\{ x^0 \}_{i=1}^n \), and denote the index-set of all variables currently labeled \(\alpha\) by \(I_{\alpha} = \{ i : x^0_i = \alpha \}\). The problem of
finding the optimal $\alpha$-$\beta$-exwap is

$$\arg\min_{x_i \in \{\alpha, x_i^0\}, x_i \in \{\alpha, \beta\}, i \notin I_\alpha} \sum_{k=1}^n E_k(x_k) + \sum_{1 \leq i < j \leq n} E_{ij}(x_i, x_j) + \sum_{1 \leq i < j < k \leq n} E_{ijk}(x_i, x_j, x_k).$$

(32)

As in the case of the expansion move for the extended energy function (Sec. 6.2), the objective function has the same structure as the one of the energy function (23) for $L = 2$ (for each variable $x_i$, associate $\alpha$ with $l_0$ and one of $x_i^0$ or $\beta$ with $l_1$, or vice versa). Therefore, it may be minimized using the minimization algorithm in [9], provided that all the $E_{ij}$ terms and all the projections of the $E_{ijk}$ terms of two variables fulfill (12), where the variables in (12) and the fixed variables in the projections may assume the labels feasible by an $\alpha$-$\beta$-exwap from the current labeling. For each variable $x_i$ these feasible labels are $\{\alpha, x_i^0\}$ (if $i \notin I_\alpha$) or $\{\alpha, \beta\}$ (if $i \in I_\alpha$). We either associate $\alpha$ with $l_0$ for all variables or associate $\alpha$ with $l_1$ for all variables. As in the case of the expansion move, $x_i^0$, $x_j^0$, and $x_k^0$ might equal any labels $\beta$, $\gamma$, and $\delta$, respectively, and the overall minimization procedure iterates over all labels $\alpha$. Therefore, the conditions under which the proposed minimization procedure works are the same as those for the expansion move type (27).

Although the exwap move type is a strict generalization of the expansion move type, a derivation similar to the one in Sec. 5 reveals that under the expansion move-related condition (19) a local minimum of the energy function (1) with respect to expansion moves is also a local minimum with respect to exwap moves. However, this relation between these two types of local minima does not hold for the extended energy function (23), which suggests that better minimization might be obtained for this energy function by using the exwap moves instead of the expansion moves.

8 Conclusion

Alternative derivations to the problems of finding the optimal swap and expansion moves were given in a short and simple manner by incorporating as a “black box” the original solution by Greig et al. [5] to the problem of optimizing a quadratic function of binary variables. By using the original solution by Greig et al. as a black box, the given construction is short and purely algebraic and is thus simpler to understand and implement than the original one in [3]. Moreover, the solution here was derived under more general conditions than the corresponding original ones. In addition, it was shown that under the swap move-related condition a local minimum with respect to expansion moves is actually also a local minimum with respect to exwap moves. All these results were extended for energy functions consisting of terms that are functions of up to three variables. This was accomplished by using as a black box Kolmogorov et al.’s algorithm in [9] for minimizing such energy functions of binary variables. The exwap move type, a generalization of the expansion and the swap move types, was introduced and an algorithm for minimizing with respect to it was derived.

We note that the derivations for the expansion moves in [8]–[9] have similarities with the corresponding ones here. Nevertheless, the former for the extended class of energy functions was case-specific and no general conditions of applicability as those in (27) were obtained. We also note that during the last stage of preparation of this paper, the swap move-related condition (16) for the class of energy functions (1) was mentioned in [10] to have been shown in [9]. However, it seems that it is not mentioned there.
References