RC25552 (WAT1508-011) August 7, 2015
Mathematics

IBM Research Report

On the Speed of Convergence of *gen*-OMP Algorithm under RIP Conditions

Aurélie Lozano, Tomasz Nowicki, Grzegorz Swirszcz
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598 USA

ON THE SPEED OF CONVERGENCE OF *gen*-OMP ALGORITHM UNDER RIP CONDITIONS

AURÉLIE LOZANO, TOMASZ NOWICKI AND GRZEGORZ ŚWIRSZCZ

ABSTRACT. Following the program of translation of algorithmic problems into a dynamical system setup, we investigate the bounds for a *gen*-OMP algorithm in the terms of iterates of a piecewise affine one dimensional map. OMP algorithms are greedy methods for sparse signal recovery or approximation from random measurements. They originate from the signal-processing community and have also been popular in the domains of function approximation, statistics and machine learning.

1. INTRODUCTION

The aim of this paper is to represent a problem of number of iterates in a *gen*-OMP (*generalized Orthogonal Matching Pursuit*) algorithm in the language of dynamical systems and to analyze it using geometrical intuitions derived from the understanding of the underlying dynamics. The class of Orthogonal Matching Pursuit techniques has received considerable attention in machine learning, statistics and signal processing as a key tool for sparse signal recovery and sparse model estimation from noisy data [11, 5, 10, 13, 14]. OMP procedures are iterative. At each step the column of the data matrix which is most correlated with the current residuals is picked. This column is then added into the set of selected columns. The residuals are then updated by projecting the observations onto the linear subspace spanned by the columns that have been selected so far and the algorithm then iterates. The representation of algorithms in dynamics proved to be successful in many cases. In [7, 8] dynamics in a functional space helped establish a tight estimation of a parameter in an algorithm describing maximal matching in graphs; in [9] a study of convergence to a neutral attractor clarified the behaviour in a routing algorithm; in [3] a study of an dynamical evolution in the space of distributions provided a better value of a critical parameter in a scheduling problem; and several papers such as [1, 2] about Error Diffusion have applications in half toning and color printing. It is our belief that both algorithmic and dynamics communities can benefit from such inter-disciplinary approach.

This work was inspired by an excellent but involved article [12]. We aimed at making the arguments more accessible and geometric, and also obtained sharper results. While the improvement would hopefully be of interest to the algorithmic community, our goal is also to expose this type of problems to the wider range of recipients, such as but not limited to dynamical systems experts.

In this section, we introduce key notation and definitions, and review the *gen*-OMP procedure. We then split the remainder of this paper into two parts. In the first part (Section 2) we deal with dynamical systems generated by a family g of one dimensional piecewise affine functions, parameterized by finite sequences of points related to the discontinuities. We prove Theorem 1.2 bounding from above the number of iterates of the function g for which the rightmost point is mapped below the leftmost discontinuity.

In the second part (Section 3) we show how to use this Theorem to estimate the number of steps of the *gen*-OMP algorithm that are sufficient to run before stopping the procedure. We define the stopping rule below. The dependence between the functions g and *gen*-OMP is described in Theorem 3.1. Finally in Theorem 1.3 we state explicitly the number of steps in *gen*-OMP.

1.1. **Piece-wise linear functions f and g.** In the definitions of both families f and g we use two constants $0 < \rho < 1 < \mu$. Let $\mathbf{Q} = (Q_M, Q_{M-1}, \dots, Q_0)$ be a finite sequence of real numbers with $Q_M = 0$ for which we denote $|\mathbf{Q}| = M$. We artificially define $Q_{-1} = \mu Q_0$. For $0 \le j < i \le M$ we define the functions

$$f_{i,j}(q) = f_{\mathbf{Q},i,j}(q) = Q_i + \left(1 - \frac{\rho}{i-j}\right)(q - Q_i) \quad \text{if} \quad q \in \left[\frac{Q_{j-1}}{\mu}, \frac{Q_j}{\mu}\right) \quad \text{and} \quad q \ge Q_i$$

and $f_{\mathbf{Q},i,j}(q) = q$ otherwise. Let

$$f(q) = f_{\mathbf{Q},\mu,\rho}(q) = \min_{0<j<i\le M} f_{\mathbf{Q},i,j}(q).$$

Define also $h_j(q) = Q_j$ if $Q_j < q < Q_{j-1}/\mu$ and $h_j(q) = q$ otherwise. Let $h(q) = \min_{0<j<M} h_j(q)$. If $Q_j \ge Q_{j-1}/\mu$ we have $h_j(q) \equiv q$. Finally define

$$g(q) = g_{\mathbf{Q},\mu,\rho}(q) = h \circ f(q)\,. \tag{1.1}$$

When it does not lead to confusion we will be writing $f_{\mathbf{Q}}$, $f_{i,j}$, $g_{\mathbf{Q}}$ etc. instead of $f_{\mathbf{Q},i,j}$, $g_{\mathbf{Q},\mu,\rho}$.

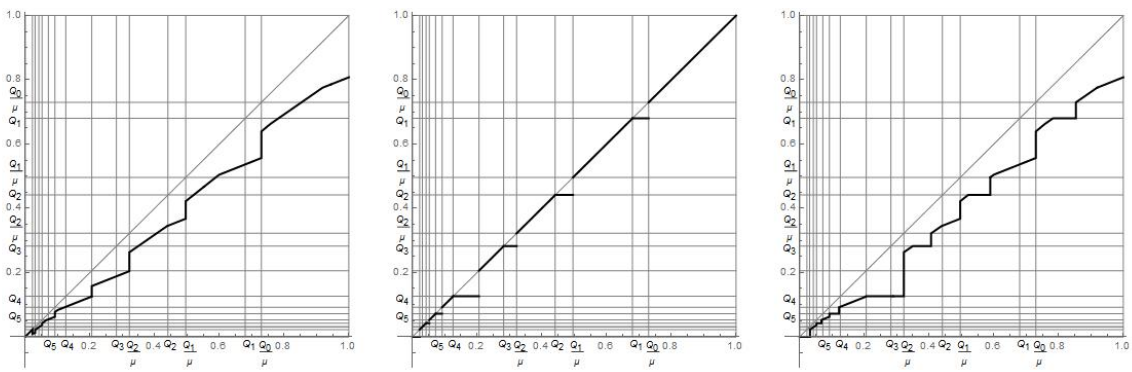

FIGURE 1. Graphs of $f_{\mathbf{Q}}$, $h_{\mathbf{Q}}$ and $g_{\mathbf{Q}}$

As usual one defines the iterates ϕ^n of a function ϕ as by $\phi^0 = \mathrm{id}$ and $\phi^{n+1} = \phi \circ \phi^n$.

We will be interested in estimating from above the number n of iterates of the function $g_{\mathbf{Q}}$ after which $g^n_{\mathbf{Q}}(\mathbf{Q}_0) = 0$. As we shall see later, this corresponds to *gen*-OMP algorithm terminating successfully. Note that $g^n_{\mathbf{Q}}(\mathbf{Q}_0) = 0$ is indeed equivalent to $g^n_{\mathbf{Q}}(Q_0) \le Q_{M-1}/\mu$, thanks to the introduction of a somewhat mysterious function h in the definition of g. Once we start discussing *gen*-OMP algorithm in more detail in Section 3.1 the reason for the presence of this function will become clear.

Remark 1.1. The condition $Q_M = 0$ can be achieved by change of coordinates (translation) $q \mapsto q - Q_M$, $Q_i \mapsto Q_i - Q_M$. Without it the definitions of $f_{\mathbf{Q},i,j}$ would be more cumbersome: the intervals of definition would be given by $(Q_j - Q_M)/\mu \le q < (Q_{j-1} - Q_M)/\mu$, and similar complications would arise in the formulas for functions h_j. Additional rescaling allows to make $Q_0 = 1$, but that does not make the formulas much simpler, so we are not using it. The artificial Q_{-1} helps defining the functions $f_{i,0}$ acting near Q_0.

Set

$$\begin{aligned}\kappa_M &= \frac{M}{\rho}\left(2\ln\mu+3\rho+3\ln 2\right),\\ \kappa_{\mathbf{Q}} &= \inf\{\kappa : \forall_{n\in\mathbb{N},n>\kappa}\ g^n(Q_0)=0\}.\end{aligned}$$

Theorem 1.2. *If $n\in\mathbb{N}$ satisfies $n\geq\kappa_{|\mathbf{Q}|}$ then $g^n(Q_0)=0$.*

In other words, for every $\mathbf{Q}$ there holds $\kappa_{\mathbf{Q}}\leq\kappa_{|\mathbf{Q}|}$.
The proof of Theorem 1.2 is given in Section 2, where we reduce it to the monotone sequences $\mathbf{Q}$.

1.2. **The steps in the *gen*-OMP algorithm.** The *gen*-OMP algorithm is used to approximate a minimum of a function $\mathcal{Q}:\mathbb{R}^p\to\mathbb{R}$ which is convex and has somewhat restricted geometry. To be blunt, it is a generalization of a quadratic function $\mathcal{Q}(x)=||Ax-y||^2$, whose minimization produces the regression coefficients x. This models the *response* vector y by a linear combination of columns of the matrix A. The columns are usually called *features* and the rows *examples*. Take a regression model of a medical treatment with (not too many) p patients and d (very large) observable variables (temperature, weight, height, partial pressures, concentration of various chemicals, medications), where y is a variable we desire to predict or control such as the expected number of days to recovery. One would prefer to have as few variables x as possible to consider in order to make a good estimate of the outcome y in this otherwise largely over-determined situation. The price to pay is the accuracy of the prediction and the cost and time of the calculation, expressed for example in the number of steps of an algorithm to be performed.

The classic OMP algorithm that can be thought of as an "iterated linear regression" seeks the approximation point with a small (compared to p) number of non-zero coordinates, by adding the coordinates one by one. Denote by $\operatorname{supp}(x)=\{i:x^i\neq 0\}$ the set of indices of non-zero coordinates of x, and $||x||_0=\operatorname{cardsupp}(x)$. Denote by $\mathcal{X}(F)$ the minimizer of $\mathcal{Q}$ on the set of coordinates F, $\mathcal{Q}(\mathcal{X}(F))\leq\mathcal{Q}(x)$ for all x with $\operatorname{supp}(x)\subset F$. Suppose that a set F of features (columns of A) is already chosen and $\mathcal{X}(F)$ was found. With the OMP algorithm we look then for the next column A_i which best aligns with the residuum $A\mathcal{X}(F)-y$, that is for which $|(Ax-y)\cdot A_i|^2/|A_i^2|$ is maximal. The next set of features will be $F'=F\cup\{i\}$ with the minimizer $\mathcal{X}(F')$. The stopping rule needs to balance the size of the resulting set of features and the accuracy of approximation.

For a convex function $\mathcal{Q}(x)$ one can generalize OMP to *gen*-OMP as follows. We assume that for any point we can determine easily both the value of $\mathcal{Q}$ and its gradient $\nabla\mathcal{Q}$ and that the cost of finding a relative minimum over all the points with a small (relative to p) number of non-zero coordinates is bearable. In *gen*-OMP at step n, given some subset $F_n\subset\{1,\ldots,p\}$ of coordinates (the starting set at $n=0$ may be empty) we look for the minimal value $q(n)$ of $\mathcal{Q}$ over all the points with the non-zero coordinates in this set only. Then we choose the largest component (over all the coordinates) of the gradient at this point, add its index to our subset, forming F_{n+1} and find the next relative minimal value $q(n+1)$. We can stop the algorithm when the relative minimal value is satisfactory, for example when the gradient is small, which suggests a closeness to the global minimum. Following [12] we define an estimate on the gradient called *Restricted Gradient Optimal Constant* as:

$$\epsilon_s(x)=\min\{\epsilon:|\nabla\mathcal{Q}(x)u|\leq\epsilon||u||_2\text{ for all }u\in\mathbb{R}^d\text{ with }||u||_0\leq s\}\,. \tag{1.2}$$

Small $\nabla(x)$ implies small $\epsilon_s(x)$.

If there exists an unknown point with M non-zero coordinates which approximates the minimum in a satisfactory way we would like to know that using *gen*-OMP we can find a point with not many more coordinates at which the value of $\mathcal{Q}$ is not much worse.

We define positive numbers $\rho_s^\pm$ as estimates of sorts of the Hessian, that is by the *Restricted Strong Convexity* condition:

$$\rho_s^- \|y-x\|_2^2 \le |\mathcal{Q}(y) - \mathcal{Q}(x) - \nabla\mathcal{Q}(x)(y-x)| \le \rho_s^+ \|y-x\|_2^2 . \tag{1.3}$$

for all $\|y-x\|_0 \le s$. This condition is closely related to the highly celebrated RIP conditions from [4]. Given S we set

$$\rho = \rho(S) = \frac{\rho_S^-}{\rho_1^+} . \tag{1.4}$$

Speaking about *gen*-OMP we introduce yet another constant c fixed throughout the paper, which in turn helps us to define μ. For any fixed $c > 0$ we define:

$$\mu = \mu(s,m) = (c+2)(c+2\frac{\rho_m^+}{\rho_s^-}) . \tag{1.5}$$

We recall that $\kappa(M,\mu,\rho) = \frac{M}{\rho}\,(2\ln\mu + 3\rho + 3\ln 2)$.

Theorem 1.3 (Number of steps in *gen*-OMP)**.** *For any $\bar{x} \in \mathbb{R}^d$ with $\bar{F} = \operatorname{supp}(\bar{x})$ and $F_0 \subset \{1,\dots,d\}$ with $M = \operatorname{card}(\operatorname{supp}(\bar{x}) \setminus F_0)$:*

If there exists an integer S satisfying:

$$S \ge |\bar{F} \cup F_0| + \kappa(M, \mu(S,M), \rho(S))$$

then for any $\kappa \ge \kappa(M,\mu(S,M),\rho(S))$ steps of the gen-*OMP algorithm starting with initial set F_0 of features we have:*

$$q(\kappa) \le \mathcal{Q}(\bar{x}) + \frac{\epsilon_S(\bar{x})^2}{c\rho_S^-} .$$

In Section 3.5, using the results from Section 3.1 we reduce this Theorem to Theorem 1.2.

Remark 1.4. The implicit assumption on S follows the setup from [12].

Remark 1.5. Zhang's result [12] states that for $\kappa_Z = 4\frac{M}{\rho(S)}\ln\frac{20\rho_M^+}{\rho_S^-}$ if there exists an $S \ge |\bar{F} \cup F_0| + \kappa_Z$ then for any $\kappa \ge \kappa_Z$ we have $q(\kappa) \le \mathcal{Q}(\bar{x}) + \frac{2.5\epsilon_S(\bar{x})^2}{\rho_S^-}$.

With $c = 0.4$ to keep the constant at ϵ the same, our result essentially improves the constant $4\ln\mu$ to $2\ln\mu$ and the constant 20 to 15 (for small $\rho(S)$).

2. Proof of Theorem 1.2

2.1. **Preliminaries.** We start the section with the following simple observation

Lemma 2.1. *Let κ be such that for every $n > \kappa$ and for every decreasing $\mathbf{Q}$ with $|\mathbf{Q}| = M$ there holds $g^n_{\mathbf{Q}}(Q_0) \le Q_M$. Then $g^n_{\mathbf{Q}}(Q_0) \le Q_M$ for every $\mathbf{Q}$ with $|\mathbf{Q}| = M$ (not necessarily decreasing).*

Proof. Let $\mathbf{Q} = (Q_M, Q_{M-1}, \dots, Q_{+1}j, Q_j, Q_{j-1}, \dots, Q_0)$ and let $\mathbf{R} = (Q_M, Q_{M-1}, \dots, Q_{j+1}, \min\{Q_j, Q_{j-1}\}, Q_{j-1}, \dots, Q_0)$ for some $0 < j < M$. If $Q_j \le Q_{j-1}$ then trivially $f_{\mathbf{Q},\mu,\rho}(q) \equiv f_{\mathbf{R},\mu,\rho}(q)$. If $Q_j > Q_{j-1}$ then for every i, q there holds

(1) $f_{\mathbf{Q},i,j}(q) \le f_{\mathbf{Q},i,j-1}(q)$ and $f_{\mathbf{R},i,j}(q) = f_{\mathbf{R},i,j-1}(q)$,
(2) $f_{\mathbf{Q},j-1,k}(q) \le f_{\mathbf{Q},j,k}(q)$ and $f_{\mathbf{R},j-1,k}(q) = f_{\mathbf{R},j,k}(q)$

therefore

$$f_{\mathbf{Q},\mu,\rho}(q) = \min_{0<k<i\le M} f_{\mathbf{Q},i,k}(q) = \min_{\substack{0<k<i\le M \\ k\ne j-1, i\ne j}} f_{\mathbf{Q},i,k}(q) = \min_{\substack{0<k<i\le M \\ k\ne j-1, i\ne j}} f_{\mathbf{R},i,k}(q) = \min_{0<k<i\le M} f_{\mathbf{R},i,k}(q),$$

thus $g_{\mathbf{Q}} \equiv g_{\mathbf{R}}$ and Lemma 2.1 follows. □

Therefore we need only to consider sequences $\mathbf{Q}$ such that $= Q_M \leq Q_{M-1} \leq \cdots \leq Q_0$ and from this point on the monotonicity of $\mathbf{Q}$ will be assumed.

We want to estimate the number of iterates $\kappa = \kappa_{\mathbf{Q}}$, such that $f^\kappa(Q_0) \leq \frac{Q_{M-1}}{\mu}$. We shall split the monotone trajectory $f^j(Q_0)$ into several parts separated by a monotone sequence of points P_t. Define K to be such that $\mu^{-K} < Q_{M-1}/Q_0 \leq \mu^{-(K-1)}$. For any t set $P_t = Q_{M-1} \cdot \mu^{K-t}$. We have $P_0 > Q_0$, $P_1 \in \left(\frac{Q_0}{\mu}, Q_0\right]$, $P_K = Q_{M-1}$ and $P_{K+1} = \frac{Q_{M-1}}{\mu}$. The trajectory between two points P_{i-1} and P_i will be split into two parts by a point $P_{i-\alpha}$ for some appropriate $\alpha \in (0,1)$. The following three Lemmata will help us to define this number.

2.2. Looking for α.

Lemma 2.2 (ξ). *For every $0 < \rho < 1$, $\mu > 1$ there exists a $\xi \in (0, \mu]$ such that for any $\gamma \in I = [1 + e^{-\rho}\xi, 1 + \xi)$ we have $\gamma^3/(\gamma - 1)^2 < 8$.*

Proof. Define $\Xi(\gamma) = \gamma^3/(\gamma-1)^2$, We look for such $\xi \leq \mu$ that the maximum of Ξ on the interval I is minimal. As Ξ is convex for $\gamma > 1$ its maximum is achieved at the endpoints of I. The value at this maximum is minimal when we choose ξ such that the values at the endpoints are equal. The positive solution to $\Xi(1 + e^{-\rho}\xi) = \Xi(1 + \xi)$ is at $\xi(\rho) = (1 + e^{-\rho/3})e^{2\rho/3}$ with the value $\Xi_{\max} = (1 + e^{-\rho/3} + e^{-2\rho/3})^3/(e^{-2\rho/3}(1 + e^{-\rho/3})^2)$ which is not larger than $7.33... < 8$, the value for $\rho = 1$. The maximal value of ξ over all ρ is at $\rho = 1$ and then $\xi_{\max} = e^{2/3} + e^{1/3} = 3.433...$. The choice of $\xi = \xi(\rho)$ satisfies the bound when $\mu \geq \xi_{\max}$ for any value of $0 < \rho < 1$. For $\mu \leq \xi_{\max}$ we can choose $\xi = \mu$. Because $\mu > 1 > e^{1/3}(e^{1/3} - 1) > e^{\rho/3}(e^{\rho/3} - 1)$ we have $\Xi(1+\mu) \geq \Xi(1+e^{-\rho}\mu)$, that is the maximum $\Xi(\gamma)$ on I is at $\Xi(1+\mu)$. But for $1 < \mu \leq \xi_{\max} < 3.5$ we have $\Xi(1 + \mu) < 7.43... < 8$. □

Remark 2.3. For *gen*-OMP usage there is no need of the second part of the proof as in *gen*-OMP we have $\mu \geq 4$.

In order to avoid rounding errors we prove that we can find integers in some intervals.

Lemma 2.4 (κ). *Given $\mu > 1$ and $0 < \rho < 1$ for any $0 < \xi \leq \mu$ there exist:*

(1) *$\overline{\gamma} \in [1 + \xi e^{-\rho}, 1 + \xi)$ with $\overline{\kappa} = \frac{\ln \frac{\mu}{\overline{\gamma}-1}}{\rho} \in \mathbb{N}$,*
(2) *$\underline{\gamma} \in [\overline{\gamma}, \overline{\gamma}e^\rho)$ with $\underline{\kappa} = \frac{\ln \underline{\gamma}}{\rho} \in \mathbb{N}$.*

Proof. The interval $[1 + \xi e^{-\rho}, 1 + \xi)$ is mapped by $\gamma \mapsto \rho^{-1} \ln \frac{\mu}{\gamma-1}$ onto $\left(\rho^{-1}\ln(\mu/\xi), \rho^{-1}\ln(\mu/\xi) + 1\right]$ of length 1, containing an integer $\overline{\kappa} > 0$, as $\mu > \xi$. Its pre-image defines $\overline{\gamma}$. Similarly the map $\gamma \mapsto \rho^{-1}\ln\gamma$ transforms $[\overline{\gamma}, e^\rho\overline{\gamma})$ onto $\left[\rho^{-1}\ln\overline{\gamma}, \rho^{-1}\ln\overline{\gamma} + 1\right)$ containing an integer $\underline{\kappa} > 0$ as $\overline{\gamma} > 1$. Its pre-image defines $\underline{\gamma} \geq \overline{\gamma}$. □

Given ρ and μ we use ξ from Lemma 2.2 to set up $\overline{\gamma}$ from Lemma 2.5 and to define:

$$\alpha = \frac{\ln((\mu - 1)\overline{\gamma} + 1)}{\ln\mu} - 1, \tag{2.1}$$

Lemma 2.5. *We have: $\overline{\gamma} = \frac{\mu^{\alpha+1}-1}{\mu-1}$, $\frac{\mu^\alpha - 1}{\mu-1} = \frac{\overline{\gamma}-1}{\mu}$ and $\alpha \in (0,1)$. Moreover $\mu^\alpha < \overline{\gamma}$.*

Proof. The equalities follow from a simple calculation and they imply the last inequality. From $(\mu-1)\overline{\gamma}+1 = \mu+(\mu-1)(\overline{\gamma}-1) > \mu$ we get $\alpha > 0$ and from $(\mu-1)\overline{\gamma}+1 = \mu^2-(\mu-1)(\mu-(\overline{\gamma}-1)) < \mu^2$ (as $\mu \geq \xi > \overline{\gamma} - 1$) we get $\alpha < 1$. □

Lemma 2.6. *With the above notation for any n, i we have:*

$$
\begin{aligned}
(1-\frac{\rho}{n})^{n\overline{\kappa}} &\leq e^{-\ln\frac{\mu}{\overline{\gamma}-1}} = \frac{\overline{\gamma}-1}{\mu} = \frac{\mu^{\alpha}-1}{\mu-1} = \frac{P_{i-\alpha}-P_i}{P_{i-1}-P_i}\,,\\
(1-\frac{\rho}{n})^{n\underline{\kappa}} &\leq e^{-\ln\underline{\gamma}} = \frac{1}{\underline{\gamma}} \leq \frac{1}{\overline{\gamma}} = \frac{\mu-1}{\mu^{\alpha+1}-1} = \frac{P_i-P_{i+1}}{P_{i-\alpha}-P_{i+1}}\,.
\end{aligned}
$$

Proof. Straightforward from the definitions of α and P_t. We used $(1-x) \leq e^{-x}$. □

For $i \in \{0,\ldots,K+1\}$ let the integer j_i be minimal such that: $Q_{j_i} \leq P_i$. By definition of K at the beginning of this subsection we have $j_0 = 0$, $j_K = M-1$ and $j_{K+1} = M$.

Using $\overline{\kappa}$ and $\underline{\kappa}$ from Lemma 2.4 we define κ_i: for $i \in \{2,\ldots,K\}$: $\kappa_i = \overline{\kappa}_i + \underline{\kappa}_i$, where $\overline{\kappa}_i = (j_i - j_{i-2})\overline{\kappa}$, $\underline{\kappa}_i = (j_{i+1} - j_{i-2})\underline{\kappa}$. For $i = 1$ we set $\overline{\kappa}_1 = j_1\overline{\kappa}$ and $\underline{\kappa}_1 = j_2\underline{\kappa}$, which follows the general definition if we set artificially $j_{-1} = 0$. For $i = K+1$: $\kappa_{K+1} = \lceil \frac{M-j_{K-1}}{\rho} \ln\mu \rceil$.

Lemma 2.7 (Bounds on κ_i)**.** *For any $1 \leq i \leq K+1$: $f^{\kappa_i}(P_{i-1}) \leq P_i$.*

Proof. Let $\Gamma_j = [Q_j/\mu, Q_{j-1}/\mu)$, the interval where $f_{i,j}(q) \neq q$. For $i = 1, \ldots K$ we have $Q_{j_i} \leq P_i < Q_{j_i-1}$, that is, after dividing by μ: $P_{i+1} \in \Gamma_{j_i}$. As $P_1 > Q_0/\mu$ we also have $P_1 \in \Gamma_0 = \Gamma_{j_0}$. In the following we shall use that for $Q < P < x, 0 \leq r \leq 1 : Q + r(x-Q) \leq P + r(x-P)$. Fix $1 \leq i \leq K$, then using Lemma 2.6 and the definition of P_t:

$$
\begin{aligned}
f^{\overline{\kappa}_i}(P_{i-1}) &\leq f^{\overline{\kappa}_i}_{j_i,j_{i-2}}(P_{i-1}) = Q_{j_i} + (1-\frac{\rho}{j_i - j_{i-2}})^{\overline{\kappa}_i}(P_{i-1} - Q_{j_i})\\
&\leq P_i + \frac{P_{i-1}-P_i}{P_{i-\alpha}-P_i}(P_{i-1}-P_i) = P_{i-\alpha}\\
f^{\underline{\kappa}_i}(P_{i-\alpha}) &\leq f^{\underline{\kappa}_i}_{j_i,j_{i-2}}(P_{i-\alpha}) = Q_{j_{i+1}} + (1-\frac{\rho}{j_{i+1} - j_{i-2}})^{\underline{\kappa}_i}(P_{i-\alpha} - Q_{j_{i+1}})\\
&\leq P_{i+1} + \frac{P_i-P_{i+1}}{P_{i-\alpha}-P_{i+1}}(P_{i-\alpha}-P_{i+1}) = P_i
\end{aligned}
$$

For $i = 1$ the argument uses artificial $j_{-1} = 0$ and that $P_0 \in \Gamma_0 = [\frac{Q_0}{\mu}, Q_0)$. Finally for $i = K+1$ we have $Q_{M-1} = P_K \in \Gamma_{j_{K-1}}$ and:

$$
\begin{aligned}
f^{\kappa_{K+1}}(P_K) &\leq f^{\kappa_{K+1}}_{M,j_{K-1}}(P_K) = \left(1-\frac{\rho}{M-j_{K-1}}\right)^{\kappa_{K+1}} P_K\\
&\leq \left(1-\frac{\rho}{M-j_{K-1}}\right)^{\frac{M-j_{K-1}}{\rho}\ln\mu} P_K < \frac{P_K}{\mu} = \frac{Q_{M-1}}{\mu} = P_{K+1}\,.
\end{aligned}
$$

□

2.3. **Completing the proof of Theorem 1.2.**

Proposition 2.8 (Total number of iterations)**.** *For $\kappa = \sum_{i=1}^{K+1} \kappa_i$ we have $\kappa \leq \kappa_M = \frac{M}{\rho}\,(2\ln\mu + 3\rho + 3\ln 2)$ and $f^{\kappa}(Q_0) \leq Q_{M-1}/\mu$.*

Proof. The estimate for $f^\kappa(Q_0)$ follows inductively from the previous Lemma. For the estimate on κ remind that we chose the parameter $0 < \xi \le \mu$ from Lemma 2.2. We have:

$$\begin{aligned}
\kappa &= \sum_{i=1}^{K+1} \kappa_i = \sum_{i=1}^{K} \kappa_i + \kappa_{K+1} = \sum_{i=1}^{K} (\overline{\kappa}_i + \underline{\kappa}_i) + \kappa_{K+1} \\
&\le (j_K + j_{K-1} - j_0 - j_{-1})\overline{\kappa} + (j_{K+1} + j_K + j_{K-1} - j_1 - j_0 - j_{-1})\underline{\kappa} + \kappa_{K+1} \\
&\le \rho^{-1}\big(2M \ln \mu + (2M + j_{K-1}) \ln \underline{\gamma} - (M + j_{K-1}) \ln(\overline{\gamma} - 1)\big) + 1 \\
&\le \rho^{-1}\left(2M \ln \mu + M \ln \frac{\underline{\gamma}^3}{(\overline{\gamma}-1)^2}\right) + 1 \le 1 + M\rho^{-1}\left(2 \ln \mu + 3\rho + \ln \frac{\overline{\gamma}^3}{(\overline{\gamma}-1)^2}\right) \\
&\le 1 + M\rho^{-1}\left(2 \ln \mu + 3\rho + 3 \ln 2\right) ,
\end{aligned}$$

where last estimates follows from Lemma 2.2. ☐

From Proposition 2.8 Theorem 1.2 follows.

3. The iterative estimation in *gen*-OMP

We make the description of the *gen*-OMP algorithm more specific. We begin by explicating how the features are selected by *gen*-OMP . Then we describe the trade-off between the number of feature selected and the quality of the solution. Next we describe and provide a set of results on what we call the "feature gaining mechanism" of *gen*-OMP . Finally, we conclude the section with the proof of Theorem 3.5.

3.1. **The *gen*-OMP algorithm.** For $x \in \mathbb{R}^d$ let $x_i \in \mathbb{R}$ denote its i^{th} coordinate and for a set of indices (features) $F \subset \{1, \dots, d\}$ let $x_{|F} \in \mathbb{R}^d$ be the restriction of x to the support F defined by:

$$(x_{|F})_i = \begin{cases} x_i & i \in F\,, \\ 0 & i \notin F\,. \end{cases}$$

We shall use some properties of the restriction such as $||x||_2^2 = ||x_{|A}||_2^2 + ||x_{|B}||_2^2 - ||x_{|A\cap B}||_2^2$. By $\operatorname{supp}(x) \subset \{1, \dots, d\}$ we denote the support of x or the set of its (non-zero) features:

$$\operatorname{supp}(x) = \{i \in \{1, \dots, d\} : x_i \neq 0\}\,.$$

The cardinality of this set will be denoted by $||x||_0 = \operatorname{card} \operatorname{supp}(x)$. Given the set of features F we denote "the minimal x":

$$\mathcal{X}(F) \text{ to be any } x \in \mathbb{R}^d \text{ with } \operatorname{supp}(x) \subset F \text{ such that } \mathcal{Q}(x) \le \mathcal{Q}(y) \text{ for all } y \text{ with } \operatorname{supp}(y) \subset F\,.$$

We note that it is possible that $\operatorname{supp}(\mathcal{X}(F)) \neq F$.

Given a point x optimal on the set of features F, $\mathcal{Q}(x) = \mathcal{Q}(\mathcal{X}(F))$ we denote by $\text{next}(F)$ and $\text{next}(x)$ respectively the set of features and any point generated by the *gen*-OMP algorithm in the following way. Denoting the partial derivative with respect to j-th variable by $\nabla_j \mathcal{Q} = \nabla \mathcal{Q} \cdot \mathbf{e}_j$ we first we choose any "maximal gradient" index j, *i.e.* any j such that $|\nabla_j \mathcal{Q}(x)| \ge |\nabla_i \mathcal{Q}(x)|$ for all $1 \le i \le d$. We set

$$\begin{aligned}
\text{next}(F) &= F \cup \{j\} \text{ and} \\
\text{next}(x) &= \mathcal{X}(\text{next}(F)).
\end{aligned}$$

Because $F \subset \text{next}(F)$ we have $\mathcal{Q}(\text{next}(x)) < \mathcal{Q}(x)$ unless $\nabla \mathcal{Q}(x) = 0$ and $x = \text{next}(x)$ is the global minimum of $\mathcal{Q}$. Starting at some set of features $F = F(0)$ and the point $x = x(0) = \mathcal{X}(F_0)$ we shall denote the iterates of the *gen*-OMP algorithm by $F(k)$ and $x(k)$ respectively thus $F(k+1) = \text{next}(F(k))$ and $x(k+1) = \text{next}(x(k))$. We shall denote the the value $\mathcal{Q}(x(k))$ by $q(k)$. The algorithm stops when some approximation condition is met.

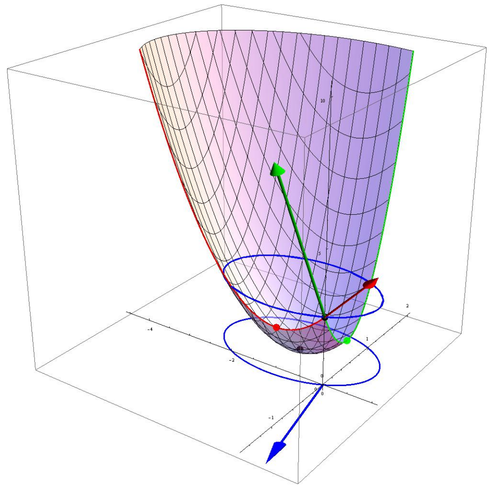

FIGURE 2. *gen*-OMP algorithm

3.2. **Approximation of a suspected almost optimal point.** The trade-off of number of features versus the closeness to the optimum can be expressed as follows.

For any point X with m non-zero coordinates whose closeness to the minimum is measured by some estimate of the gradient of $\mathcal{Q}$ at X we will estimate the number of steps from any initial point $x(0)$ until we reach the value which is comparable with this estimate. It turns out that the *gen*-OMP algorithm either decreases the value $q(k) = \mathcal{Q}(x(k))$ by a controllable amount or adds the coordinates of X to the coordinates of $x(k)$. After a number of steps k which depends on $||X||_0$, but not on d nor on p, either (a) $F(k)$ contains $E = \operatorname{supp}(X)$, but not much more, in which case $q(k) \le \mathcal{Q}(X)$ and $\operatorname{card} F(k) \le \operatorname{const} \operatorname{card}(E)$ or (b) $q(k)$ will be close enough to $Q = \mathcal{Q}(X)$ and $x(k)$ will be close to X, both of them in terms of the gradient estimate (1.3).

Remark 3.1. In the model function $\mathcal{Q}(x) = ||Ax - y||_2^2 = (Ax - y)^T (Ax - y)$, $x \in \mathbb{R}^d$, $y \in \mathbb{R}^p$, $A \in \mathbb{R}^{p \times d}$, the difference in the 1.3 formula is equal to $(y - x)^T A^T A (y - x)$, and the constants $\rho_s^{\pm}$ can be estimated respectively by the minimum and maximum of the set of all eigenvalues of the matrices $A_F^T A_F$, where F runs over all subsets of $\{1, \ldots, d\}$ with cardinality at most s and A_F is formed by the columns of A with indices in F.

Remark 3.2. The function $s \mapsto \rho_s^{\pm}$ is non-increasing, in particular for any m, s we have $\rho_m^+ \ge \rho_s^-$.

3.3. **The iterative model.** Suppose that for a set of features F we found the optimal point $x = \mathcal{X}(F)$:

$$\mathcal{Q}(\mathcal{X}(F)) \le \mathcal{Q}(z) \text{ for any } z \text{ with } \operatorname{supp}(z) \subset F\,.$$

When we apply *gen*-OMP we get the coordinate j of maximal gradient and the next minimum $\operatorname{next}(x) = \mathcal{X}(F \cup \{j\}) = \mathcal{X}(\operatorname{next}(F))$ relative to the new set of features. We estimate $\mathcal{Q}(\operatorname{next}(x)) \le \min_\alpha \mathcal{Q}(x + \alpha \mathbf{e}_j)$, and using (1.3):

$$\mathcal{Q}(\operatorname{next}(x)) \le \min_\alpha \left(\mathcal{Q}(x) + \alpha \nabla \mathcal{Q}(x)(e_j) + \rho_1^+ \alpha^2 \right) = \mathcal{Q}(x) - \frac{(\nabla_j \mathcal{Q}(x))^2}{4\rho_1^+}\,,$$

where the minimum was achieved at $\alpha = -\nabla_j \mathcal{Q}(x)/2\rho_1^+$.

By *gen*-OMP choice $|\nabla_j \mathcal{Q}(x)|$ was maximal and by optimality $\nabla_i \mathcal{Q}(x) = 0$ for $i \in F$. For any point z with features $E = \operatorname{supp}(z)$ we have

$$\begin{aligned}
|\nabla_j \mathcal{Q}(x)| \sum_{i \in E \setminus F} |z_i| &= \sum_{i \in E \setminus F} |\nabla_j \mathcal{Q}(x)||z_i - x_i| \geq \sum_{i \in E \setminus F} |\nabla_i \mathcal{Q}(x)||z_i - x_i| \\
&= \sum_{i \in E \cup F} |\nabla_i \mathcal{Q}(x)||z_i - x_i| \geq |\sum_{i \in E \cup F} \nabla_i \mathcal{Q}(x)(z_i - x_i)| \\
&= |\nabla \mathcal{Q}(x)(z - x)|,
\end{aligned}$$

where we used $x_i = 0$ for $i \notin F \supset \operatorname{supp}(x)$, and similarly $z_i = 0$ for $i \notin E (= \operatorname{supp}(z))$. Again using (1.3) and setting $s = \operatorname{card} F \cup E$ we continue:

$$|\nabla_j \mathcal{Q}(x)| \sum_{i \in E \setminus F} |z_i| \geq |\nabla \mathcal{Q}(z)(z - x)| \geq -\nabla \mathcal{Q}(x)(z - x) \geq \mathcal{Q}(x) - \mathcal{Q}(z) + \rho_s^- ||z - x||_2^2,$$

It follows that for any point z with $\mathcal{Q}(z) \leq \mathcal{Q}(x)$ we have

$$\mathcal{Q}(\operatorname{next}(x)) \leq \mathcal{Q}(x) - \frac{(\mathcal{Q}(x) - \mathcal{Q}(z) + \rho{-}s||x - z||^2)^2}{4\rho_1^+ (\sum_{i \in E \setminus F} |z_i|)^2}.$$

Consider the function $\phi(\xi) = \xi - (\xi + \alpha)^2/4\beta$, with $\beta > \alpha > 0$ (if $\beta < \alpha$ then $\phi(\xi) < 0$). The line $\lambda(\xi)$, tangent to ϕ at α lies above ϕ, hence $\phi(\xi) \leq \lambda(\xi) = (1 - \alpha/\beta)\xi$. (The same can be achieved via $(\alpha + \beta)^2 \geq 4\alpha\beta$.) It follows that:

$$\mathcal{Q}(\operatorname{next}(x)) \leq \mathcal{Q}(z) + \left(1 - \frac{\rho_s^-}{\rho_1^+} \frac{||x - z||_2^2}{(\sum_{i \in E \setminus F} |z_i|)^2}\right) (\mathcal{Q}(x) - \mathcal{Q}(z)).$$

We have

$$||x - z||_2^2 \geq \sum_{i \in E \setminus F} |z_i|^2 \geq \frac{(\sum_i |z_i|)^2}{\operatorname{card} E \setminus F},$$

and finally with $\rho = \frac{\rho_s^-}{\rho_1^+}$ and $l = \operatorname{card} E \setminus F$:

$$\mathcal{Q}(\operatorname{next}(x)) \leq \mathcal{Q}(z) + \left(1 - \frac{\rho}{l}\right) (\mathcal{Q}(x) - \mathcal{Q}(z)). \tag{3.1}$$

We shall use this inequality with different z's in order to estimate the decrease of the *gen*-OMP sequence. Remark that the equality may happen at each iteration and that the estimate depends only on the values of the function $\mathcal{Q}$. This is the prototype of the functions $f_{\mathbf{Q},i,j}$.

3.4. **The Feature Gaining Mechanism.** From now on we shall fix a point X with $||X||_0 = M$. We do not assume anything else about this point, in particular we do not know its coordinates.

After some (minimal) number of steps K the iterates of the *gen*-OMP algorithm starting at a point $x(0)$ would reach $q(K) = \mathcal{Q}(x(K)) < \mathcal{Q}(X) + \text{const}\, \epsilon_S^2(X)$, where $S = ||X - x(k)||_0 \leq k_0 + K + M$, and $\epsilon_S(X)$ was defined in (1.2). It is possible as $\mathcal{Q}(x_K)$ decreases to the absolute minimum and ϵ_s increases with K. The heuristic is that M is small (compared to d and p) and that $\epsilon(X)$ is small, so that X is close to the minimum, with a small number of features. The number K represents the number of steps to achieve a relative closeness to X and hence also to the minimum, it is our hope that K depends linearly on M and so we would find a satisfactory approximation of the minimum with a number of coordinates comparable with the number of features of the (unknown) candidate X, and not with a possibly large number of examples p or even a very large total number of features d.

First we show under which conditions the set $F(k)$ gains the features (indices) from the set $\mathrm{supp}\,(X)$. We need two technical definitions. For any $L \leq ||X||_0$ we define:

$$k_L = \min\{k : \mathrm{card}\,(F(k) \cap \mathrm{supp}\,(X)) \geq L\}, \tag{3.2}$$

$$E_L = E_L(X,k) = \mathrm{supp}\,(F(k_L)) \cap \mathrm{supp}\,(X). \tag{3.3}$$

Note that such k_L and E_L exist always as $\mathrm{card}\,F(k) = k$, and that by definition $\mathrm{card}\,E_L = L$. The feature gaining sequence E_L was constructed by adding the features of X in the order in which they appeared in the *gen*-OMP sequence.

Proposition 3.3 (Feature Gaining Mechanism)**.** *With the above notation if* $||x(k) - X||_2 < ||X - X_{|E_L}||_2$*, then* $k \geq k_{L+1}$*. In other words* $\mathrm{card}\,(F(k) \cap \mathrm{supp}\,(X)) \geq L+1$.

Proof. If $\mathrm{card}\,(F(k)) \cap \mathrm{supp}\,(X)) > L$ we are done. Otherwise we have $F(k) \cap \mathrm{supp}\,(X) \subset E_L$ and $\mathrm{supp}\,(X) \setminus F(k) \supset \mathrm{supp}\,(X) \setminus E_L$

$$\begin{aligned} ||x(k) - X||_2^2 &= ||(x(k) - X)_{|F(k)} - X_{|\mathrm{supp}\,(X)\setminus F(k)}||_2^2 \\ &= ||(x(k) - X)_{|F(k)}||_2^2 + ||X_{|\mathrm{supp}\,(X)\setminus F(k)}||_2^2 \\ &\geq ||X_{|\mathrm{supp}\,(X)\setminus E_L}||_2^2 = ||X - X_{|E_L}||_2^2\,, \end{aligned}$$

contradicting the assumption. □

Remark 3.4. The same proof of Proposition 3.3 works with any p-norm.

Remark 3.5. In Proposition 3.3 one can consider any sequence, not necessarily generated by *gen*-OMP , $x(k)$ satisfying $\mathrm{card}\,(\mathrm{supp}\,(x(k+1)) \geq \mathrm{card}\,(\mathrm{supp}\,(x(k)) + 1$, with $F(k) = \mathrm{supp}\,(x(k))$.

We shall use Proposition 3.9 on the *gen*-OMP sequence in order to provide the inequality in the assumption of Proposition 3.3 to gain the features from the point X into the features of the *gen*-OMP points.

Remark 3.6. It is possible to construct different feature gaining sequences, for example one can choose the sequence of largest coordinates of X.

The following Lemmata lead to Proposition 3.9 which is crucial to establish the condition in the Feature Gaining Mechanism.

Lemma 3.7. *Let* a, b *be two points with* $s = ||a - b||_0$*. If for* $\epsilon_s(a)$ *from* (1.2) *and some constants* $c, \rho > 0$ *we have:* $\mathcal{Q}(b) - \mathcal{Q}(a) \geq \frac{\epsilon_s(a)^2}{c\rho}$ *then:*

$$||b - a||_2^2 \leq (\mathcal{Q}(b) - \mathcal{Q}(a)) \frac{c\rho + 2\rho_s^-}{(\rho_s^-)^2} \tag{3.4}$$

$$||b - a||_2^2 \geq (\mathcal{Q}(b) - \mathcal{Q}(a)) \frac{1}{c\rho + 2\rho_s^+}\,. \tag{3.5}$$

Remark 3.8. In case $a \neq b$ sharp inequality $\mathcal{Q}(b) - \mathcal{Q}(a) > \frac{\epsilon_s(a)^2}{c\rho}$ yields sharp inequalities (3.4) and (3.5).

Proof. Inequality (3.4):

$$\begin{aligned} \mathcal{Q}(b) - \mathcal{Q}(a) &\geq \rho_s^- ||b - a||_2^2 + \nabla \mathcal{Q}(a)(b - a) \text{ by (1.3)} \\ &\geq \rho_s^- ||b - a||_2^2 - \epsilon_s(a) ||b - a||_2 \text{ by (1.2)}. \end{aligned}$$

Hence $\rho_s^-||b-a||_2^2 - \epsilon_s(a)||b-a||_2 - (\mathcal{Q}(b)-\mathcal{Q}(a)) \leq 0$ and therefore:

$$||b-a||_2 \leq \frac{\epsilon_s(a) + \sqrt{\epsilon_s(a)^2 + 4\rho_s^-\,(\mathcal{Q}(b)-\mathcal{Q}(a))}}{2\rho_s^-} \leq \frac{\sqrt{\epsilon_s(a)^2 + 2\rho_s^-\,(\mathcal{Q}(b)-\mathcal{Q}(a))}}{\rho_s^-}\,,$$

where the last inequality follows from concavity of $\sqrt{\cdot}$. But by assumption, $\epsilon_s(a)^2 \leq c\rho\,(\mathcal{Q}(b)-\mathcal{Q}(a))$, so:

$$||b-a||_2 \leq \frac{\sqrt{(c\rho + 2\rho_s^-)}}{\rho_s^-}\sqrt{\mathcal{Q}(b)-\mathcal{Q}(a)}\,.$$

By inequality (3.5) we have

$$\begin{aligned}\rho_s^+||b-a||_2^2 &\geq \mathcal{Q}(b)-\mathcal{Q}(a) - \nabla\mathcal{Q}(a)(b-a) \text{ by (1.3)}\\ &\geq \mathcal{Q}(b)-\mathcal{Q}(a) - \epsilon_s(a)||b-a|| \text{ by (1.2)}.\end{aligned}$$

Similarly as before that means that $\rho_s^+||b-a||_2^2 + \epsilon_s(a)||b-a|| - (\mathcal{Q}(b)-\mathcal{Q}(a)) \geq 0$ and therefore, as $\mathcal{Q}(b) \geq \mathcal{Q}(a)$ by assumption and $||b-a| \geq 0$ by definition, we have:

$$\begin{aligned}||b-a||_2 &\geq \frac{-\epsilon_s(a) + \sqrt{\epsilon_s(a)^2 + 4\rho_s^+\,(\mathcal{Q}(b)-\mathcal{Q}(a))}}{2\rho_s^+}\\ &= \frac{2\,(\mathcal{Q}(b)-\mathcal{Q}(a))}{\epsilon_s(a) + \sqrt{\epsilon_s(a)^2 + 4\rho_s^+\,(\mathcal{Q}(b)-\mathcal{Q}(a))}}\\ &\geq \frac{\mathcal{Q}(b)-\mathcal{Q}(a)}{\sqrt{\epsilon_s(a)^2 + 2\rho_s^+\,(\mathcal{Q}(b)-\mathcal{Q}(a))}}\,,\end{aligned}$$

where again we used concavity in the last inequality. From $\epsilon_s(a)^2 \leq c\rho\,(\mathcal{Q}(b)-\mathcal{Q}(a))$, we finally get:

$$||b-a||_2 \geq \frac{\sqrt{\mathcal{Q}(b)-\mathcal{Q}(a)}}{\sqrt{c\rho + 2\rho_s^+}}\,.$$

□

Next Proposition shows the condition for the Feature Gaining Mechanism to work.

Proposition 3.9. *Let u, y, z be three points with $||y-u||_0 = s$ and $||z-u||_0 = m$. If for some $c > 0$ and $\rho \leq \rho_s^-$:*

$$\frac{\epsilon_s^2(u)}{c\rho} \leq \mathcal{Q}(y)-\mathcal{Q}(u) \leq \frac{1}{(c+2)(c+2\frac{\rho_m^+}{\rho_s^-})}\,(\mathcal{Q}(z)-\mathcal{Q}(u))$$

then

$$||y-u||_2 < ||z-u||_2\,.$$

Proof. Note that as $\rho_m^+ \geq \rho_s^-$ we have $\frac{1}{(c+2)(c+2\rho_m^+/\rho_s^-)} < 1/4 < 1$ which means that both u, y and u, z fulfil the assumptions of Lemma 3.7. We use (3.4) with $b = y$ and $a = u$ and (3.5) with $b = z$

and $a = u$ (with a sharp inequality in this case) and get:

$$\begin{aligned}
||y-u||_2^2 &\leq (\mathcal{Q}(y)-\mathcal{Q}(u))\,\frac{c\rho+2\rho_s^-}{(\rho_s^-)^2} \\
\text{by assumption} \quad &\leq \frac{1}{(c+2)(c+2\frac{\rho_m^+}{\rho_s^-})}\,(\mathcal{Q}(z)-\mathcal{Q}(u))\,\frac{c\rho+2\rho_s^-}{(\rho_s^-)^2} \\
&< \frac{1}{(c+2)(c+2\frac{\rho_m^+}{\rho_s^-})}\,\frac{c\rho+2\rho_s^-}{(\rho_s^-)^2}(c\rho+2\rho_m^+)||z-u||_2^2 \\
&= \frac{(c\rho+2\rho_s^-)(c\rho+2\rho_m^+)}{(c\rho_s^-+2\rho_s^-)(c\rho_s^-+2\rho_m^+)}||z-u||_2^2 \leq ||z-u||_2^2\,.
\end{aligned}$$

□

Remark 3.10. $\frac{1}{c+2}\frac{1}{c+2\frac{\rho_m^+}{\rho_s^-}} \geq \frac{1}{(c+2)^2}\frac{\rho_s^-}{\rho_m^+}$ which for $c=\frac{2}{5}$ yields $\frac{25}{144}\frac{\rho_s^-}{\rho_m^+} > \frac{1}{6}\frac{\rho_s^-}{\rho_m^+}$.

Corollary 3.11. *If for some k and m, S there holds $||x(k)-X||_0 \leq S$, $L \leq m$ and*

$$\mathcal{Q}(x(k))-\mathcal{Q}(X) \leq \frac{1}{c+2}\frac{1}{c+2\frac{\rho_m^+}{\rho_S^-}}\left(\mathcal{Q}(X_{|E_L})-\mathcal{Q}(X)\right)\,,$$

then either $\epsilon_S(X)^2 > c\rho_S^-\,(\mathcal{Q}(x(l))-\mathcal{Q}(X))$ or $\mathrm{card}\,(F(k)\cap \mathrm{supp}\,(X)) > L$.

Proof. If the condition on $\epsilon_S(X)$ is not satisfied then we can use Proposition 3.9 with $y = x(l)$, $u = X$, $z = X_{|E_L}$ in order to fulfill the assumption of Proposition 3.3 and use Proposition 3.3. □

3.5. **Proof of Theorem 1.3.** As in the beginning of subsection 3.4 let X be fixed with M and K defined there. From now on we shall assume that for the *gen*-OMP sequence starting at $x(0)$, $x(k+1) = \mathrm{next}(x(k))$, $k = 0,\ldots,S$ the following condition is satisfied $\mathcal{Q}(x(k))-\mathcal{Q}(X) > \frac{\epsilon_S(X)^2}{c\rho_S^-}$, as otherwise for some $k < S$ we have the statement of the Theorem.
Define μ, k_L and E_L as in (1.5), (3.2) and (3.3). Denote $Q_L = \mathcal{Q}(X_{|E_L})-\mathcal{Q}(X)$ and recall

$$f_{i,j}(q) = \begin{cases} Q_i + (1-\frac{\rho}{i-j})(Q_i - q) & Q_i \leq q < \frac{Q_{j-1}}{\mu}\,, \\ q & q \text{ otherwise}\,. \end{cases}$$

For any $0 \leq x(k) < \frac{Q_j}{\mu}$ we have $\mathrm{card}\,(\mathrm{supp}\,(X_{|E_i}) \setminus \mathrm{supp}\,(x(k))) < i-j$ and thus by (3.1) we have $q(k+1)-\mathcal{Q}(X) = \mathcal{Q}(\mathrm{next}(x(k)))-\mathcal{Q}(X) \leq Q_i + (1-\frac{\rho}{i-j})\,[\mathcal{Q}(x(k))-\mathcal{Q}(X)-Q_i] = f_{i,j}(q(k)-\mathcal{Q}(X))$. Moreover, by Corollary 3.11, if for any L $q(k+1)-\mathcal{Q}(X) < \frac{Q_L}{\mu}$ then $q(k+1)-\mathcal{Q}(X) \leq Q_{L+1}$. Thus $q(k+1)-\mathcal{Q}(X) \leq h(f_{i,j}(q(k)-\mathcal{Q}(X)))$. Therefore the number of iterates of the *gen*-OMP algorithm can be estimated from above by the number of iterates of the function g (1.1) at $q = q(k)-\mathcal{Q}(X)$, starting at $q(0)-\mathcal{Q}(X)$ and ending with $q(K)-\mathcal{Q}(X) \leq Q_{M-1}/\mu$ because then all the features of X will be caught and hence $\mathcal{Q}(x(K)) \leq \mathcal{Q}(X)$. Thus, the problem of estimating the number of steps of *gen*-OMP has been reduced to the following:
Given μ and ρ - what is the maximal number of iterates of g from $q = Q_0$ to $q \leq Q_{M-1}/\mu$, where the maximality is searched over all the sequences $\mathbf{Q}$? But that was established by Theorem 1.2.

References

1. R. Adler, T. Nowicki, G. Świrszcz, C. Tresser, and S. Winograd, *Error diffusion on simplices: Invariant regions, tessellations and acuteness.* Preprint, (2012).
2. R. Adler, T. Nowicki, G. Świrszcz and C. Tresser. Convex dynamics with constant input. *Erg. Th. and Dyn. Sys.*, 30(4):957-972, (2010).
3. M. Bienkowski, J. Byrka, M. Chrobak, N. Dobbs, T. Nowicki, M. Sviridenko, G. Świrszcz, N. E. Young: Approximation Algorithms for the Joint Replenishment Problem with Deadlines. *ICALP* 1:135-147 (2013)
4. E.J. Candès, T. Tao: Decoding by linear programming. *IEEE Trans. on Inf. Th.*, 51:4203-4215 (2005)
5. D. Donoho, *Compressed sensing*, IEEE Trans. Info. Theory **52** (2006), no. 4, 1289–1306.
6. S. Foucart A note on guaranteed sparse recovery via ℓ_1-minimization. *Applied and Computational Harmonic Analysis* 29.1:97-103, (2010)
7. D. Gamarnik, T. Nowicki, G. Świrszcz: Maximum weight independent sets and matchings in sparse random graphs. Exact results using the local weak convergence method. *Random Struct. Algorithms* 28(1):76-106 (2006)
8. ———: Invariant probability measures and dynamics of exponential linear type maps. *Erg. Th. and Dyn. Sys.*, 28(1):1479-1495, (2008)
9. T. Nowicki, M. Sviridenko, G. Swirszcz, S. Winograd On the rate of convergence to the neutral attractor of a family of one-dimensional maps. *Fund. Math.* 206:253-269, (2009)
10. V. N. Temlyakov, *Greedy approximation*, Acta Numerica **17** (2008), 235–409.
11. J. Tropp, *Greed is good: algorithmic results for sparse approximation*, IEEE Trans. Info Theory **50** (2004), no. 10, 2231–2242.
12. T. Zhang. Sparse Recovery with Orthogonal Matching Pursuit under RIP. *IEEE Trans. Info. Th.*, 57:5215-6221, (2011).
13. ———, *Adaptive forward-backward greedy algorithm for sparse learning with linear models*, Neural Information Processing Systems (NIPS) 21, 2008.
14. ———, *On the consistency of feature selection using greedy least squares regression*, Journal of Machine Learning Research **10** (2009), 555–568.