QCS-SGM+: Improved Quantized Compressed Sensing with Score-Based Generative Models

Xiangming Meng¹*, Yoshiyuki Kabashima²
¹The Zhejiang University-University of Illinois Urbana-Champaign Institute, Zhejiang University, China
²Institute for Physics of Intelligence & Department of Physics, The University of Tokyo, Japan
xiangmingmeng@intl.zju.edu.cn, kaba@phys.s.u-tokyo.ac.jp

Abstract

In practical compressed sensing (CS), the obtained measurements typically necessitate quantization to a limited number of bits prior to transmission or storage. This nonlinear quantization process poses significant recovery challenges, particularly with extreme coarse quantization such as 1-bit. Recently, an efficient algorithm called QCS-SGM was proposed for quantized CS (QCS) which utilizes score-based generative models (SGM) as an implicit prior. Due to the adeptness of SGM in capturing the intricate structures of natural signals, QCS-SGM substantially outperforms previous QCS methods. However, QCS-SGM is constrained to (approximately) row-orthogonal sensing matrices as the computation of the likelihood score becomes intractable otherwise. To address this limitation, we introduce an advanced variant of QCS-SGM, termed QCS-SGM+, capable of handling general matrices effectively. The key idea is a Bayesian inference perspective on the likelihood score computation, wherein expectation propagation is employed for its approximate computation. Extensive experiments are conducted, demonstrating the substantial superiority of QCS-SGM+ over QCS-SGM for general sensing matrices beyond mere row-orthogonality.

Introduction

Compressed sensing (CS) has emerged as a ubiquitous paradigm in signal processing and machine learning, aiming to accurately reconstruct high-dimensional signals from a limited number of measurements (Donoho 2006; Candès and Wakin 2008). The success of CS hinges on the assumption that, while the target signal may be high-dimensional, it possesses an inherent low-dimensional representation such as sparsity or low-rankness. Conventional CS models typically assume direct access to analog (continuous) measurements with infinite precision. In practice, however, analog measurements must be quantized into a finite number of digital bits using an analog-to-digital converter (ADC or quantizer) before further transmission, storage, or processing can occur (Boufounos and Baraniuk 2008; Zymnis, Boyd, and Candes 2009; Dai and Milenkovic 2011; Plan and Vershynin 2012, 2013; Jacques et al. 2013; Xu and Kabashima 2013; Xu, Kabashima, and Zdeborová 2014; Awasthi et al. 2016; Meng, Wu, and Zhu 2018; Jung et al. 2021; Liu et al. 2020; Liu and Liu 2022; Zhu et al. 2022; Meng and Kabashima 2023b). Among these, the recently proposed algorithm QCS-SGM (Meng and Kabashima 2023b) demonstrates exceptional state-of-the-art (SOTA) reconstruction performance under low-precision quantization levels. The key idea of QCS-SGM lies in utilizing the powerful score-based generative models (SGM, also known as diffusion models) (Song and Ermon 2019, 2020; Sohl-Dickstein et al. 2015; Ho, Jain, and Abbeel 2020; Nichol and Dhariwal 2021) as an implicit prior for the target signal. Intuitively, from a Bayesian perspective, the more accurate the prior obtained, the fewer observations required. Owing to SGM’s ability to capture the intricate structure of the target signal, QCS-SGM can accurately reconstruct even from a small number of severely quantized noisy measurements.

While QCS-SGM exhibits remarkable performance, it has one fundamental limitation: it is derived under the assumption that the sensing matrix \( \mathbf{A} \) is (approximately) row-orthogonal. For general matrices, QCS-SGM’s performance will deteriorate since its computation of the likelihood score (defined in (4)) becomes less accurate (Meng and Kabashima 2023b). Although row-orthogonal matrices are prevalent in conventional CS, in practical applications one might encounter other types of matrices due to non-ideal physical constraints or design choices. In fact, the investigation of general sensing matrices has long been an active and important topic, such as the popular ill-conditioned matrices and correlated matrices, among others (Manoel et al. 2015;...

*Corresponding author.

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.
Schniter, Rangan, and Fletcher 2016; Tanaka 2018; Ihara et al. 2018; Rangan, Schniter, and Fletcher 2019; Venkataramanan, Kögler, and Mondelli 2022; Zhu et al. 2022; Fan 2022). Despite these advances, much of existing work concentrates on standard CS or QCS with handcrafted sparsity. The study of QCS using SGM (diffusion models) for general sensing matrices, however, still remains a largely untouched research area. This paper try to address this problem and the main contributions are summarized as follows:

- We propose an advanced variant of QCS-SGM, designated as QCS-SGM+, which addresses the inherent limitation of the original QCS-SGM. Specifically, by treating the likelihood score computation as a Bayesian inference problem, QCS-SGM+ utilizes the well-established expectation propagation (EP) algorithm (Minka 2001) to yield a more refined approximation of the otherwise intractable likelihood score for general matrices.

- We validate the effectiveness of the proposed QCS-SGM+ in various experimental settings, encompassing diverse real-world datasets, distinct general matrices, and different noise levels. In each scenario, QCS-SGM+ consistently demonstrates remarkably superior performance over existing methods.

Related Work

Quantized compressed sensing (QCS) was first introduced in Boufounos and Baraniuk (2008); Zymnis, Boyd, and Candes (2009), after which it has become an important research topic and various QCS algorithms have been proposed, including some theoretical analysis (Dai and Milenkovic 2011; Plan and Vershynin 2012, 2013; Jacques et al. 2013; Xu and Kabashima 2013; Xu, Kabashima, and Zdeborová 2014; Awasthi et al. 2016; Meng, Wu, and Zhu 2018; Jung et al. 2021; Liu et al. 2020; Liu and Liu 2022; Zhu et al. 2022). With the recent advent of deep generative models (Goodfellow et al. 2014; Kingma and Welling 2013; Rezende and Mohamed 2015; Song and Ermon 2019, 2020; Sohl-Dickstein et al. 2015; Ho, Jain, and Abbeel 2020; Nichol and Dhariwal 2021), there has been a rising interest in CS methods with data-driven priors (Bora et al. 2017; Hand and Joshi 2019; Asim et al. 2020; Pan et al. 2021; Meng and Kabashima 2022). Specifically, following the pioneering CSGM framework (Bora et al. 2017), the prior $p(x)$ of $x$ is learned through a generative model, such as VAE (Kingma and Welling 2013), GAN (Goodfellow et al. 2014), and score-based generative models (SGM) or diffusion models (DM) (Song and Ermon 2019, 2020; Sohl-Dickstein et al. 2015; Ho, Jain, and Abbeel 2020; Nichol and Dhariwal 2021). In the case of QCS, Liu et al. (2020); Liu and Liu (2022) extended the CSGM framework to nonlinear observations such as 1-bit CS using VAE and GAN (in particular DCGAN (Radford, Metz, and Chintala 2015)). Surprisingly, SGM or DM (Song and Ermon 2019, 2020; Ho, Jain, and Abbeel 2020; Nichol and Dhariwal 2021) have demonstrated superior effectiveness, even surpassing state-of-the-art GAN (Goodfellow et al. 2014) and VAE (Kingma and Welling 2013) in generating diverse natural sources. In line with this, Meng and Kabashima (2023b) recently proposed a novel algorithm, QCS-SGM, employing SGM as an implicit prior, achieving state-of-the-art reconstruction performances for QCS. However, its application remains confined to (approximate) row-orthogonal sensing matrices.

Preliminary

System Model

The problem of quantized CS (QCS) can be mathematically formulated as follows (Boufounos and Baraniuk 2008; Zymnis, Boyd, and Candes 2009)

$$y = Q(Ax + n),$$

where the goal is to recover an unknown high-dimensional signal $x \in \mathbb{R}^{N \times 1}$ from a set of quantized measurements $y \in \mathbb{R}^{M \times 1}$, where $A \in \mathbb{R}^{M \times N}$ is a known linear sensing matrix, $n \sim \mathcal{N}(0, \sigma^2 I)$ is an i.i.d. additive Gaussian noise, and $Q(\cdot) : \mathbb{R}^{M \times 1} \rightarrow Q^{M \times 1}$ is an element-wise quantizer function which maps each element into a finite (or countable) set of codewords $Q$, i.e., $y_{m} = Q(z_{m} + n_{m}) \in Q$, or equivalently $(z_{m} + n_{m}) \in Q^{-1}(q_{m}), m = 1, 2, ..., M$, where $z_{m}$ is the $m$-th element of $z = Ax$. Same as Meng and Kabashima (2023b), we consider the uniform quantizer with $Q$ quantization bits (resolution). The quantization code-words $Q = \{q_r \}$ consist of $2^Q$ elements, each with a quantization interval $Q^{-1}(q_r) = [l_q, u_q)$, where $l_q$ and $u_q$ are the lower and upper quantization threshold associated with the codeword $q_r$. In the extreme 1-bit case, i.e., $Q = 1$, only the signs are observed so that (1) reduces to

$$y = \text{sign}(Ax + n),$$

which corresponds to the well-known 1-bit CS and the quantization codewords are $Q = \{-1, +1\}$.

QCS-SGM: Quantized CS with SGM

Compared to standard CS without quantization, QCS is particularly more challenging due to two key factors: (1) quantization leads to information loss, especially with low quantization resolution; (2) the nonlinearity of quantization operations can cause standard CS algorithms to deteriorate when applied directly. Recently, inspired by the prowess of SCM (Song and Ermon 2019; Song et al. 2020) in density estimation, Meng and Kabashima (2023b) proposed an efficient method called QCS-SGM for QCS which can accurately reconstruct the target signal from a small number of severely quantized noisy measurements. The basic idea of QCS-SGM is to perform posterior sampling from $p(x | y)$ by using a learned SGM as an implicit prior (Meng and Kabashima 2023b). Specifically, by utilizing the annealed Langevin dynamics (ALD) (Song et al. 2020), the posterior samples can be iteratively obtained as follows: For $1 \leq t \leq T$,

$$x_t = x_{t-1} + \alpha_t \nabla_{x_t} \log p(x_{t-1} | y) + \sqrt{2\alpha_t} z_t,$$

where the conditional (posterior) score $\nabla_{x_t} \log p(x_t | y)$ is required. Using the Bayesian rule, the $\nabla_{x_t} \log p(x_t | y)$ is decomposed into two terms

$$\nabla_{x_t} \log p(x_t | y) = \nabla_{x_t} \log p(x_t) + \nabla_{x_t} \log p(y | x_t),$$

(4)
including the unconditional score $\nabla_x \log p(x_t)$ (called prior score in Meng and Kabashima (2023b)), and the conditional score $\nabla_x \log p(y | x_t)$ (called likelihood score in Meng and Kabashima (2023b)). While the prior score $\nabla_x \log p(x_t)$ can be readily computed using a pre-trained score network, the likelihood score $\nabla_x \log p(y | x_t)$ is generally intractable. To circumvent this difficulty, Meng and Kabashima (2023b) proposed a simple yet effective approximation of $\nabla_x \log p(y | x_t)$ under an uninformative prior assumption, whereby $x_t$ is approximated as $x_t = x + \beta \hat{n}_t$, where $\hat{n}_t \sim \mathcal{N}(0, I)$. As a result, substituting it into (1), we obtain an equivalent representation as

$$y = Q(AX_t + \hat{n}_t),$$

where $\beta = \beta_0 \sigma^2 I + \beta_1 \text{diag}(x_t)$ and $\beta_0 \sigma^2 I$, $\beta_1 \text{diag}(x_t)$ are a sequence of noise scales satisfying $\beta_0 \sigma^2 I > \beta_1 \text{diag}(x_t)$ (Song et al. 2020). From (5), an approximate solution for general matrices

$$\text{by EP to derive an effective factorized approximation of } f_b(\hat{n}_t) \equiv \mathcal{N}(\hat{n}_t; 0, C_t^{-1}).$$

After EP, the original coupled prior node $f_b(\hat{n}_t)$ in Figure 1 (a) is decoupled into a series of fully-factorized prior nodes $f_c(\hat{n}_{t,m})$, $m = 1...M$ in (b). Hence, a closed-form solution for $p(y | z_t = A x_t)$ can be obtained, thereby enabling us to compute the pseudo-likelihood score $\nabla_x \log p(y | z_t = A x_t)$. The details of our derivation via EP are illustrated as follows. Essentially, it approximates the partition function $p(y | z_t = A x_t)$ (7) in three different ways as follows:

$$\tilde{p}(y | z_t = A x_t) \approx \begin{cases} p(y | z_t = A x_t) \approx \frac{f_a(\hat{n}_t) \prod_{m=1}^M f_a(\hat{n}_{t,m})}{\tilde{p}(y | z_t = A x_t)} & (a) \\
\int \prod_{m=1}^M f_a(\hat{n}_{t,m}) \mathcal{N}(\hat{n}_{t,m}; \frac{h_F}{\tau_F}, \frac{1}{\tau_F}) d\hat{n}_t & (b) \\
\int \prod_{m=1}^M \mathcal{N}(\hat{n}_{t,m}; \frac{h_F}{\tau_F}, \frac{1}{\tau_F}) f_b(\hat{n}_t) d\hat{n}_t & (c) \end{cases}$$

A New Perspective

Our key insight is that, the pseudo-likelihood term $\tilde{p}(y | z_t = A x_t)$ (6) concerning $x_t$ can be alternatively interpreted as the partition function (normalization factor) of a posterior distribution concerning $\hat{n}_t$ (rather than $x_t$), i.e.,

$$f_a(\hat{n}_t) \prod_{m=1}^M f_a(\hat{n}_{t,m})$$

where $f_a(\hat{n}_t) \equiv \mathcal{N}(\hat{n}_t; 0, C_t^{-1})$ acts as the prior distribution, and $f_a(\hat{n}_{t,m}) \equiv 1 ((z_{t,m} + \hat{n}_{t,m}) \in Q^{-1}(y_m))$ acts as the likelihood distribution. As computing the partition function is one fundamental problem in Bayesian inference and various approximations have been studied, such as a perspective on $p(y | z_t = A x_t)$ (6) as partition function provides us with one solution using well-studied approximate Bayesian inference methods (Wainwright and Jordan 2008).

**Pseudo-Likelihood Score via EP**

Due to its efficacy and available theoretical guarantee for some general matrices (Oppen and Winther 2001, 2005; Takahashi and Kabashima 2020), we resort to the well-known expectation propagation (EP) (Minka 2001) (also known as moment matching (Oppen and Winther 2005)) to approximately compute the intractable partition function $p(y | z_t = A x_t)$ (7), i.e., the pseudo-likelihood $\tilde{p}(y | z_t = A x_t)$ (6). As illustrated in Figure 1, the basic idea is to apply EP to derive an effective factorized approximation of $f_b(\hat{n}_t) \equiv \mathcal{N}(\hat{n}_t; 0, C_t^{-1})$. After EP, the original coupled prior node $f_b(\hat{n}_t)$ in Figure 1 (a) is decoupled into a series of fully-factorized prior nodes $f_c(\hat{n}_{t,m})$, $m = 1...M$ in (b). Hence, a closed-form solution for $\tilde{p}(y | z_t = A x_t)$ (7) can be obtained, thereby enabling us to compute the pseudo-likelihood score $\nabla_x \log p(y | z_t = A x_t)$.

The details of our derivation via EP are illustrated as follows. Essentially, it approximates the partition function $\tilde{p}(y | z_t = A x_t)$ (7) in three different ways as follows:

$$\tilde{p}(y | z_t = A x_t) \approx \begin{cases} \frac{f_a(\hat{n}_t) \prod_{m=1}^M f_a(\hat{n}_{t,m})}{\tilde{p}(y | z_t = A x_t)} & (a) \\
\int \prod_{m=1}^M f_a(\hat{n}_{t,m}) \mathcal{N}(\hat{n}_{t,m}; h_F, \frac{1}{\tau_F}) d\hat{n}_t & (b) \\
\int \prod_{m=1}^M \mathcal{N}(\hat{n}_{t,m}; h_F, \frac{1}{\tau_F}) f_b(\hat{n}_t) d\hat{n}_t & (c) \end{cases}$$

Intuitively, (8-a) approximates the correlated multivariate Gaussian $f_b(\hat{n}_t) \equiv \mathcal{N}(\hat{n}_t; 0, C_t^{-1})$ with a product of independent Gaussians $\prod_{m=1}^M \mathcal{N}(\hat{n}_{t,m}; h_F, \frac{1}{\tau_F})$, (8-b) approximates the non-Gaussian likelihood $f_a(\hat{n}_t) \equiv \prod_{m=1}^M \mathcal{N}(\hat{n}_{t,m}; h_F, \frac{1}{\tau_F})$, and (8-c) combines the two approximations together. In contrast to the original intractable partition function $p(y | z_t = A x_t)$ (7), all the three approximations (8-a), (8-b), (8-c) become tractable, leading to three closed-form approximations to the posterior mean $\hat{\nu}_{t,m}$ and variance $\mathbb{V}[\hat{n}_{t,m}]$ of $\hat{n}_{t,m}$ w.r.t. $p(\hat{n}_t | y)$ (7), which are denoted as $(m_a, \ell_a), (m_b, \ell_b), (m_c, \ell_c)$, respectively, and can be computed as follows

$$m_a = \frac{h_F}{\tau_F} - \frac{1}{\sqrt{2\pi\tau^F}} \exp\left(-\frac{x_m^2}{2}\right) - \frac{1}{\sqrt{2\pi\tau^F}} \exp\left(-\frac{y_m^2}{2}\right),$$

$$\ell^a = \frac{1}{\sqrt{2\pi\tau^F}} \left[ \frac{\hat{n}_{t,m}^2}{\tau_F} - \frac{\hat{n}_{t,m} x_m}{\tau_F} - \frac{\hat{n}_{t,m} y_m}{\tau_F} + \frac{x_m y_m}{\tau_F} \right],$$

$$\ell^b = \frac{1}{\sqrt{2\pi\tau^F}} \left[ \frac{\hat{n}_{t,m}^2}{\tau_F} - \frac{\hat{n}_{t,m} x_m}{\tau_F} - \frac{\hat{n}_{t,m} y_m}{\tau_F} + \frac{x_m y_m}{\tau_F} \right],$$

$$\ell^c = \frac{1}{\sqrt{2\pi\tau^F}} \left[ \frac{\hat{n}_{t,m}^2}{\tau_F} - \frac{\hat{n}_{t,m} x_m}{\tau_F} - \frac{\hat{n}_{t,m} y_m}{\tau_F} + \frac{x_m y_m}{\tau_F} \right],$$

The results might differ in a scaling factor but does not affect the score function.
After EP

\[
\begin{align*}
{\mathcal{N}}(\tilde{\eta}; 0, \mathbf{C}^{-1})\end{align*}
\]

Figure 1: A schematic of the basic idea of QCS-SGM+ in computing the intractable pseudo-likelihood \(\hat{p}(y|z_t = \mathbf{A}x_t)\) for general matrices. The subscript \(t\) is dropped for simplicity. EP results in an effective factorized approximation of \(f_b(\tilde{\eta})\) so that a closed-form solution of \(\hat{p}(y|z_t = \mathbf{A}x_t)\) can be achieved.

Interestingly, the above implementation details of EP can be illustrated via iterative message passing in the corresponding factor graphs (Minka 2001; Wainwright and Jordan 2008). As shown in Figure 1, \(\lambda(\tilde{n}_m, \frac{h^F}{\sqrt{\tau^F}}, \frac{1}{\sqrt{\tau^F}})\) corresponds to the message \(m_{a\rightarrow b}(\tilde{n}_m)\) from factor node \(f_b\) to factor node \(f_a\), where \(\chi(\tilde{n}_m, \frac{h^F}{\sqrt{\tau^F}}, \frac{1}{\sqrt{\tau^F}})\) corresponds to the message \(m_{a\rightarrow b}(\tilde{n}_m)\) from factor node \(f_a\) to factor node \(f_b\). The three approximations (8-a), (8-b), and (8-c) correspond to the combined results of incoming messages in EP at factor node \(f_a\), variable node \(n_m\), factor node \(f_b\), respectively.

After EP, as shown in Figure 1 (b), we can obtain an alternative factorized representation, leading to a closed-form approximation of \(\hat{p}(y|z_t = \mathbf{A}x_t)\) as follows

\[
\hat{p}(y|z_t = \mathbf{A}x_t) \approx \frac{e^{\langle \mathbf{B}_t \mathbf{y} \rangle}}{2} \left[ \text{erfc}(-\tilde{u}_{y_m}) - \text{erfc}(-\tilde{l}_{y_m}) \right],
\]

where

\[
\begin{align*}
\tilde{u}_{y_m} &= -\sqrt{2\tau^F}z_{t,m} - \frac{h^F}{\sqrt{\tau^F}} + u_{y_m}\sqrt{\tau^F}, \\
\tilde{l}_{y_m} &= -\sqrt{2\tau^F}z_{t,m} - \frac{h^F}{\sqrt{\tau^F}} + l_{y_m}\sqrt{\tau^F}.
\end{align*}
\]

Therefore, from (6) and (20), the intractable pseudo-likelihood score \(\nabla_{x_t} \log p(y | x_t)\) under quantized measurements \(y\) in (1) can be approximated as

\[
\nabla_{x_t} \log p(y | x_t) \simeq \mathbf{A}^\top \mathbf{G} (\beta_t, y, \mathbf{A}, z_t, h^F, \tau^F),
\]

where

\[
\mathbf{G} (\beta_t, y, \mathbf{A}, z_t, h^F, \tau^F) = [g_1, g_2, ..., g_M]^\top \in \mathbb{R}^{M \times 1}
\]

with the \(m\)-th element being

\[
g_m = \frac{\sqrt{2\tau^F} \left[ \exp \left( -\frac{\tilde{u}_{y_m}^2}{2} \right) - \exp \left( -\frac{\tilde{l}_{y_m}^2}{2} \right) \right]}{\sqrt{\pi} \left[ \text{erfc}(-\tilde{u}_{y_m}/\sqrt{2}) - \text{erfc}(-\tilde{l}_{y_m}/\sqrt{2}) \right]}.
\]
Algorithm 1: QCS-SGM+

\begin{algorithm}
\label{alg:qcs-sgm+}
\textbf{Input:} \{\beta_t\}_{t=1}^{T}, \epsilon, \gamma, \text{IterEP}, K, y, A, \sigma^2, quantization thresholds \{[u_q, v_q]|q \in Q\}
\textbf{Initialization:} \mathbf{x}_0^t \sim \mathcal{U}(0,1)
\begin{algorithmic}[1]
\For {$t = 1$ to $T$}
\For {$k = 1$ to $K$}
\State Draw $x^t_k \sim \mathcal{N}(0,1)$
\EndFor
\State \textbf{Initialization:} $h^F, \tau^F, h^G, \tau^G$
\For {$i = 1$ to $\text{IterEP}$}
\State $h^G = \frac{m^x - h^F}{h^F}$
\State $\tau^G = \frac{1}{\chi^x} - \tau^F$
\State $h^F = \frac{m^b - h^G}{h^G}$
\State $\tau^F = \frac{\chi^b}{\chi^x} - \tau^G$
\EndFor
\EndFor
\State Compute $\nabla_{x^t_k} \log p(y|x^t_k)$ as (23)
\State $\mathbf{x}^t_{k+1} = \mathbf{x}^t_{k} + \alpha \left[\mathbf{g}(\mathbf{x}^t_{k-1}, \beta_t) + \gamma \nabla_{x^t_k} \log p(y|x^t_k)\right] + \sqrt{2\alpha \chi^x} \epsilon^t_k$
\State $\mathbf{x}^t_{K+1} \leftarrow \mathbf{x}^t_{K}$
\end{algorithmic}
\textbf{Output:} $\tilde{x} = \mathbf{x}^T_F$
\end{algorithm}

Resultant Algorithm: QCS-SGM+

By combining the pseudo-likelihood idea (23) approximated via EP and the prior score from SGM, we readily obtain an enhanced version of QCS-SGM, dubbed as QCS-SGM+. The details of QCS-SGM+ are shown in Algorithm 1 where the number of iterations of EP is denoted as IterEP. A scaling factor $\gamma$ is introduced, which is empirically found useful to further improve the performance.

Note that while it seems from (11, 12) that matrix inversion $(\tau^2 I + C_t)^{-1}$ is needed in each iteration for every $C_t$, this is actually not the case since there exists a single efficient implementation method using singular value decomposition (SVD) similar to Rangan, Schniter, and Fletcher (2019); Meng and Kabashima (2022). Specifically, denote $A = U \Sigma V^T$ as the SVD of $A$ and $\Sigma$ as the element-wise square of singular values, i.e., diagonal elements of $\Sigma$, then after some simple algebra, it can be shown that the terms $m^b, \chi^b$ involving a matrix inverse can be efficiently computed as follows

$$m^b = \Upsilon \text{diag} \left( \frac{\alpha^2 + \beta_2 \Sigma^2}{\tau_G (\alpha^2 + \beta_2 \Sigma^2) + 1} \right) U^T h^G,$$

$$\chi^b = \left< \frac{\alpha^2 + \beta_2 \Sigma^2}{\tau_G (\alpha^2 + \beta_2 \Sigma^2) + 1} \right>,\quad (25, 26)$$

where $\left< \cdot \right>$ is the average value of the elements in a vector. It can be seen from (25, 26) that one simply needs to replace the values of $\beta_t, \tau_G$ for different iterations. Hence, the main computational burden lies in the SVD of sensing matrix $A$, but only one time is required in the whole QCS-SGM+.

Relation to QCS-SGM

For row-orthogonal matrices $A$, the covariance matrix $C^{-1}_t = \beta^2 I + \beta_2 A A^T$ becomes diagonal and thus the prior node $f_{\epsilon}(\tilde{n}_t) \equiv N(0; C_t^{-1})$ in Figure 1 (a) already fully factorizes and thus QCS-SGM+ reduces to QCS-SGM (Meng and Kabashima 2023b). For general matrices $A$, however, there is no such equivalence even though QCS-SGM can still be applied pretending $C_t^{-1}$ to be diagonal, i.e., QCS-SGM directly diagonalizes $C_t^{-1}$ by extracting its main diagonal elements. The fundamental difference between QCS-SGM and QCS-SGM+ can be better illustrated from the new perspective on $\tilde{p}(y|x_t) = \text{partition function} \int f_{\epsilon}(\tilde{n}_t) \prod_{m=1}^M f_a(\tilde{n}_{t,m}) d\tilde{n}_t (7)$. QCS-SGM naively diagonalizes the correlated Gaussian $f_{\epsilon}(\tilde{n}_t) \equiv N(0; C_t^{-1})$, which ignores the effect of $\prod_{m=1}^M f_a(\tilde{n}_{t,m})$; by contrast, QCS-SGM+ considers $f_{\epsilon}(\tilde{n}_t) \prod_{m=1}^M f_a(\tilde{n}_{t,m})$ as a whole and diagonalizes $f_{\epsilon}(\tilde{n}_t)$ by explicitly taking into account the effect of $\prod_{m=1}^M f_a(\tilde{n}_{t,m})$, thereby leading to better performances than QCS-SGM for general matrices $A$.

Experiments

We empirically demonstrate the efficacy of the proposed QCS-SGM+ in various scenarios. The source code is available at https://github.com/mengxiangming/QCS-SGM-plus. More results can be found in the appendix of the arXiv version Meng and Kabashima (2023a).

Specifically, we investigate two popular general sensing matrices $A$ beyond row-orthogonal:

(a) ill-conditioned matrices (Rangan, Schniter, and Fletcher 2019; Schniter, Rangan, and Fletcher 2016; Venkataramanan, Kögl, and Mondelli 2022; Fan 2022):

$$A = V \Sigma U^T,$$

where $V$ and $U$ are independent Henri-distributed matrices, and the nonzero singular values of $A$ satisfy $\frac{\lambda_i}{\lambda_{i+1}} = \kappa^{1/M}$, where $\kappa$ is the condition number. Such matrices can have an arbitrary spectral distribution and often arise in practical applications (Venkataramanan, Kögl, and Mondelli 2022).

(b) correlated matrices (Shiu et al. 2000; Zhu et al. 2022): $A$ is constructed as $A = R_L H R_R$, where $L = R_L^\perp \in \mathbb{R}^{M \times M}$ and $R_R = R_R^\perp \in \mathbb{R}^{N \times N}$, the $(i, j)$th element of both $R_L$ and $R_R$ is $\rho^{|i-j|}$ and $\rho$ is termed as the correlation coefficient here. $H \in \mathbb{R}^{M \times N}$ is a random matrix whose elements are drawn i.i.d. from $\mathcal{N}(0, 1)$.

Datasets: We consider three popular datasets: MNIST (LeCun and Cortes 2010), CIFAR-10 (Krizhevsky and Hinton 2009), and CelebA (Liu et al. 2015). For CelebA dataset, we cropped each face image to a 64 × 64 RGB image, resulting in $N = 64 \times 64 \times 3 = 12288$ inputs per image.

QCS-SGM+: Same as QCS-SGM (Meng and Kabashima 2023b), we adopt the NCSNv2 (Song and Ermon 2020) in all cases. For MNIIST, the NCSNv2 was trained with a similar training set up as CIFAR-10 in Song and Ermon (2020), while for CIFAR-10, and CelebA, we use the pre-trained models from https://drive.google.com/drive/folders/1217uhIvLg9zYNKOR3XTRFSurt4miQrd.
Figure 2: Qualitative comparisons of different methods under 1-bit CS on MNIST and CelebA for ill-conditioned $A$ ($\kappa = 10^3$ for MNIST and $\kappa = 10^6$ for CelebA) when $M < N$. It can be seen that the proposed QCS-SGM+ achieves consistently better results than other methods.

Figure 3: Reconstructed images on CIFAR-10 with QCS-SGM and QCS-SGM+, respectively, under 1-3 bit CS when the condition number of $A$ is 1000, $M = 2000$, $\sigma = 0.1$. It can be seen that QCS-SGM+ is able to generate clear images even with $M < N$ 1-3 bit noisy quantized measurements, whereas the original QCS-SGM yields only vague or blurry results.

**Comparison Methods:** We compare QCS-SGM+ with not only the state-of-the-art QCS-SGM (Meng and Kabashima 2023b), but also two other leading methods before QCS-SGM, namely BIPG (Liu et al. 2020) and OneShot (Liu and Liu 2022).

**Qualitative Results**

Figure 2 shows some qualitative results of QCS-SGM+ and QCS-SGM, BIPG, and OneShot under 1-bit CS for ill-conditioned $A$. It can be seen from Figure 2 that the proposed QCS-SGM+ achieves remarkably better performances than all previous methods and can well reconstructs the target images from only a few $M < N$ sign measurements even when the sensing matrix $A$ is highly ill-conditioned. To further demonstrate the efficacy of QCS-SGM+ under different quantization resolutions (e.g., 1-3 bits), Figure 3 shows some results of QCS-SGM and QCS-SGM+, respectively, with condition number of $A$ being $\kappa = 10^3$ and $\sigma = 0.1$. It can be seen that QCS-SGM+ is able to generate clear images even with $M < N$ 1-3 bit noisy quantized measurements, whereas the original QCS-SGM yields only vague or blurry results.

**Quantitative Results**

The quantitative comparison in terms of the peak signal-to-noise ratio (PSNR) is evaluated. First, Figure 4 illustrates the PSNR results of QCS-SGM+ and QCS-SGM, in the case of 1-bit CS with MNIST, CIFAR-10, and CelebA, for ill-conditioned $A$ and correlated $A$, respectively. It can be seen that in all cases, the proposed QCS-SGM+ achieves much better performances, demonstrating the superiority of QCS-SGM+ over QCS-SGM for more general sensing matrices $A$. We also evaluate the effect of Gaussian noise $n \sim \mathcal{N}(n; 0, \sigma^2 I)$ by conducting experiments on 1-bit CS with different levels of noise standard deviation (std) $\sigma$. As shown in Figure 5, due to the potential dithering effect, QCS-SGM+ with noise can sometimes achieve slightly better results than that without noise. Generally, QCS-SGM+ is robust to noise and can achieve good results in a large range of $n$, while significantly outperforming QCS-SGM.
In this paper, we propose an improved version of Quantized Compressed Sensing with Score-Based Generative Models (QCS-SGM), termed as QCS-SGM+, in the case of general sensing matrices. By viewing the likelihood computation as a Bayesian inference problem, QCS-SGM+ approximates the intractable likelihood score using the well-known expectation propagation (EP). To verify the effectiveness of QCS-SGM+, we conducted experiments on a variety of baseline datasets, demonstrating that QCS-SGM+ significantly outperforms QCS-SGM by a large margin for general sensing matrices. There are several limitations of QCS-SGM+. For example, QCS-SGM+ requires EP message passing, which is computationally slower than QCS-SGM. Also, same as QCS-SGM, it requires a large number of iterations to generate one posterior sample. As future work, it is important to further reduce the complexity of QCS-SGM+ and develop more efficient alternatives with more advanced diffusion models. Moreover, a theoretical analysis of both QCS-SGM and QCS-SGM+ is also an important future direction.
Acknowledgements

This work was supported by NSFC No. 62306277, China, and JSPS KAKENHI Nos. 17H00764, 18K11463, and 19H01812, 22H05117, and JST CREST Grant Number JP-MJCR1912, Japan.

References


LeCun, Y.; and Cortes, C. 2010. MNIST handwritten digit database.


